NBER WORKING PAPER SERIES

IMPROVING THE MEASUREMENT OF EARNINGS DYNAMICS

Moira Daly
Dmytro Hryshko
Iourii Manovskii

Working Paper 22938
http://www.nber.org/papers/w22938

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
December 2016

We benefited from comments of participants at numerous conferences and seminars. Support from the National Science Foundation Grants No. SES-0922406, SES-1357903, and the Danish Social Science Research Council (FSE) grant No. 300279 is gratefully acknowledged. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Improving the Measurement of Earnings Dynamics
Moira Daly, Dmytro Hryshko, and Iourii Manovskii
NBER Working Paper No. 22938
December 2016
JEL No. D31,D91,E21

ABSTRACT

The stochastic process for earnings is the key element of incomplete markets models in modern quantitative macroeconomics. We show that a simple modification of the canonical process used in the literature leads to a dramatic improvement in the measurement of earnings dynamics in administrative and survey data alike. Empirically, earnings at the start or end of earnings spells are lower and more volatile than the observations in the interior of earnings histories, reflecting the effects of working less than the full year as well as deviations of wages due to e.g. tenure effects. Ignoring these properties of earnings, as is standard in the literature, leads to a substantial mismeasurement of the variances of permanent and transitory shocks and induces the large and widely documented divergence in the estimates of these variances based on fitting the earnings moments in levels or growth rates. Accounting for these effects enables more accurate analysis using quantitative models with permanent and transitory earnings risk, and improves empirical estimates of consumption insurance against permanent earnings shocks.

Moira Daly
Department of Economics
Centre for Economic and Business Research
Copenhagen Business School
Porcelaenshaven 16A, 2000
Frederiksberg
Denmark
moda.eco@cbs.dk

Dmytro Hryshko
Department of Economics
University of Alberta
8-14 HM Tory Building
Edmonton, AB, T6G2H4
Canada
dhryshko@ualberta.ca

Iourii Manovskii
Department of Economics
University of Pennsylvania
160 McNeil Building
3718 Locust Walk
Philadelphia, PA 19104
and NBER
manovski@econ.upenn.edu

1 Introduction

The central element of many models in modern quantitative macroeconomics with heterogeneous agents is either an exogenously specified or an endogenously determined stochastic process for individual earnings. For example, in the models with incomplete insurance markets, the properties of the earnings process serve as key determinants of the evolution of consumption, assets, and other economic choices over the life cycle and across individuals.[1] Following the seminal contribution by Friedman (1957), modern consumption theory recognizes that consumption should respond more to the longer-lasting or permanent than to transitory innovations in earnings. This explains the keen interest in the literature in measuring the variances of these components using the variants of the permanent/transitory earnings decomposition[2] written, in its basic form, as:

$$\begin{aligned} y_{it} &= \alpha_i + p_{it} + \tau_{it} \\ p_{it} &= \phi_p p_{it-1} + \xi_{it} \\ \tau_{it} &= \theta(L)\epsilon_{it}, \end{aligned} \tag{1}$$

where log-earnings y_{it} of individual i at time t consists of the permanent component, p_{it}, and the transitory component, τ_{it}. If ϕ_p is close to 1, the shocks ξ_{it} are highly persistent (and are truly permanent if ϕ_p is 1), and if $\theta(L) = 1$ (where $\theta(L)$ is a moving average polynomial in the lag operator L), the shocks ϵ_{it} are completely transitory.

In addition to determining equilibrium consumption and wealth distributions, the variance and persistence of the shocks ξ_{it} and ϵ_{it} have important implications for policy design. For example, they are key to determining the optimal design of the bankruptcy code in Livshits, MacGee, and Tertilt (2007), they govern the impact of the welfare system on household savings in Hubbard, Skinner, and Zeldes (1995), stimulus effects of fiscal policy in Heathcote (2005), as well as the optimal design of the tax system in Banks and Diamond (2010) and Farhi and Werning (2012). Moreover, there is great interest in understanding whether the dramatic increase in earnings dispersion over the last few decades in the U.S. is due to the increase in the variances of persistent or transitory shocks (e.g., Gottschalk and Moffitt (1994)). This understanding is relevant for determining why consumption inequality did not increase nearly as much (e.g., Krueger and Perri (2006), Blundell, Pistaferri, and Preston (2008), Heathcote, Storesletten, and Violante (2010), Attanasio, Hurst, and Pistaferri (2012)). Knowing the stochastic nature of earnings is also essential for the design of active labor market policies. For example, Meghir and Pistaferri (2011) suggest that income maintenance policies might be

[1]See, e.g., Deaton (1991), Carroll (1997), Castañeda, Díaz-Giménez, and Ríos-Rull (2003).

[2]This decomposition was pioneered by Friedman and Kuznets (1954) and found to have empirical support by MaCurdy (1982), Abowd and Card (1989), and Meghir and Pistaferri (2004), among others. A prominent alternative, e.g., Guvenen (2009), allows for less persistent shocks but individual-specific trends in earnings.

an appropriate response to changes in inequality driven by transitory shocks, while training programs are potentially more relevant to counteracting the effects of permanent shocks.

Unfortunately, despite their manifest importance, there is no consensus in the literature on the sizes of the shocks ϵ_{it} and ξ_{it}. In particular, using the same data, the estimates of the earnings process in Eq. (1) when targeting the moments of log-earnings in levels are dramatically different from the estimates obtained when fitting the moments of log-earnings in differences. Although this discrepancy was first documented using survey-based data, it remained undiminished when the focus of the literature has shifted to relying more on administrative datasets.[3] These datasets are typically orders of magnitude larger than survey-based ones; free of sampling issues; do not suffer from the typical issues of attrition; are based on administrative sources, such as tax records; and are considered highly reliable and free of issues of systematic non-response or measurement errors that typically plague survey-based data. Yet, despite numerous attractive properties, these datasets must also have features that lead to the large discrepancy in the estimates based on moments in growth rates and in levels.

Such observations led Heathcote, Perri, and Violante (2010) to conclude that the widely used model of earnings dynamics in Eq. (1) is misspecified. However, the nature of this potential misspecification is unknown. This challenges our confidence in the conclusions of the models that incorporate this earnings process. Even if this misspecified process is used as a model input due to the lack of a better alternative, it is unclear whether it is more appropriate to parameterize it using the estimates targeting the moments in levels or in differences in the data. Relatedly, in the literature that endogenizes the earnings process,[4] it is unclear whether the process implied by the model should be compared to the one estimated in the data using the specification in levels or in differences, given that estimating the reduced-form process (1) on the model-generated data does not give rise to the observed discrepancy.

In this paper, we uncover an important source of this misspecification. Estimation of the parameters of the earnings process in the literature is based on fitting the entire set of autocovariance moments for levels or differences of log-earnings. However, even when estimation is based on the same set of observations in the data, computation of the autocovariance moments in levels and differences is effectively based on different information. To clarify with an extreme example, consider an individual with a single earnings observation in the sample. This observation will contribute to the estimated variance of earnings in levels, but it will not contribute to any moment in differences. More generally, earnings observations adjacent to a missing one (e.g., observations at the start or at the end of individual's earnings history) also contribute differently to moments in levels and differences. If earnings observations surrounding the missing ones were random draws from the rest of earnings histories, this would not

[3] Recent contributions include Blundell, Graber, and Mogstad (2015), DeBacker, Heim, Panousi, Ramnath, and Vidangos (2013), Domeij and Flodén (2010), Guvenen, Ozcan, and Song (2014), among others.

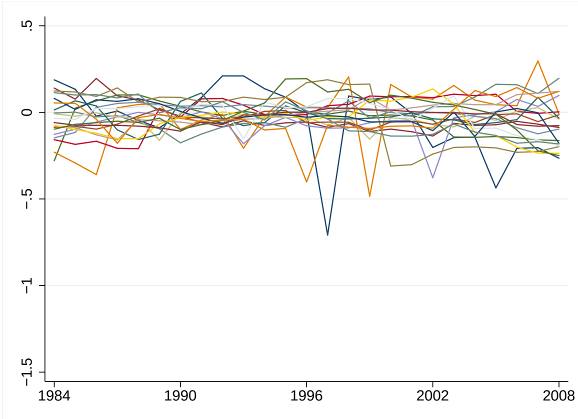
[4] E.g., Huggett, Ventura, and Yaron (2011) and Postel-Vinay and Turon (2010).

matter. However, in the data these observations are much lower than the typical ones and more volatile. We will show formally that this raises the variance of transitory shocks when estimation relies on the moments in levels and the variance of permanent shocks recovered by estimation based on the moments in differences.

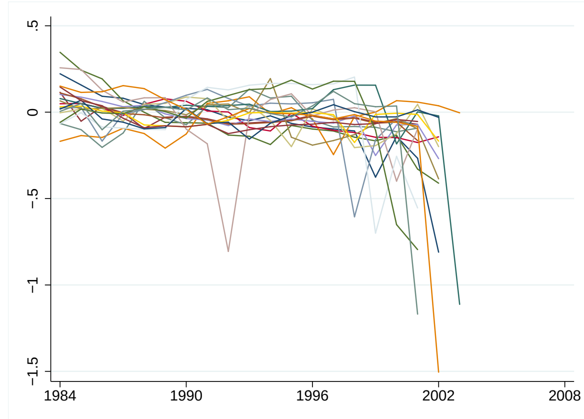
In the first set of quantitative experiments in the paper we assess the magnitude of these effects using large administrative datasets from Denmark and Germany. The Danish data contain complete earnings histories of each resident of Denmark from 1981 through 2006. The German data are a 2% random sample of social security numbers. For these individuals, the complete earnings history from 1975 through 2008 is available. These samples are sufficiently large to allow analysis at the level of particular age cohorts, making it possible to focus on a parsimonious earnings model in (1), sidestepping the issue of modeling cohort effects. Moreover, the large size of the data enables reliable estimation when replicating the design of samples typically used in the literature. Specifically, we consider a balanced sample spanning 25 (26) years in German (Danish) data, a sample with 9 or more consecutive observations, as in e.g., Browning, Ejrnæs, and Alvarez (2010) and Meghir and Pistaferri (2004), and a sample with 20 or more not necessarily consecutive observations as in e.g., Guvenen (2009). Our smallest Danish sample is comprised of about 67,000 individuals and 1.7 million observations, while our smallest German sample contains about 10,000 individuals with more than 200,000 observations.

Using the unbalanced samples in both datasets, we find, consistent with the literature, a substantially higher estimated variance of permanent (transitory) shocks targeting the moments of earnings in growth rates (levels). In contrast, we find that the discrepancy is nearly absent in balanced samples drawn from the two datasets. To highlight the differences between the earnings trajectories in balanced and unbalanced samples that we argue induce these results, in Figure 1 we plot 20 random (residual) earnings paths for four subsamples in the German data over 1984-2008 period.[5] Panel (a) depicts earnings paths for individuals in the balanced sample. For the vast majority of these individuals, their first year in the sample does not coincide with the first year of their earnings history. Similarly, their last year in the sample mechanically truncates earnings histories, implying that it is not the last year of the earnings spell of individuals in the sample. Thus, the mean and the variance of earnings in the first and the last sample years are similar to those in the other years. This stands in sharp contrast to earnings histories of individuals entering and/or exiting the data in the interior of the sample window. For example, panel (b) plots earnings paths for individuals leaving the sample early – in 2002–2004. Panel (c) plots earnings paths for individuals entering the sample late – in 1989–1991. Finally, panel (d) plots earnings paths for individuals within an incomplete spell that starts later than 1984 and ends earlier than 2008. Clearly, the earnings

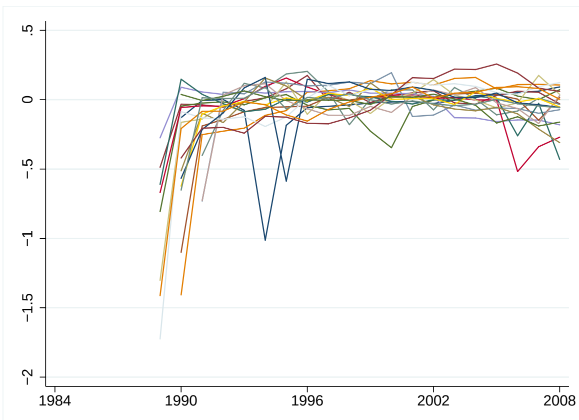
[5] For this figure, we used data for individuals whose mean residual earnings belong to the 45th to 55th percentiles of the distribution of the mean individual residual earnings.



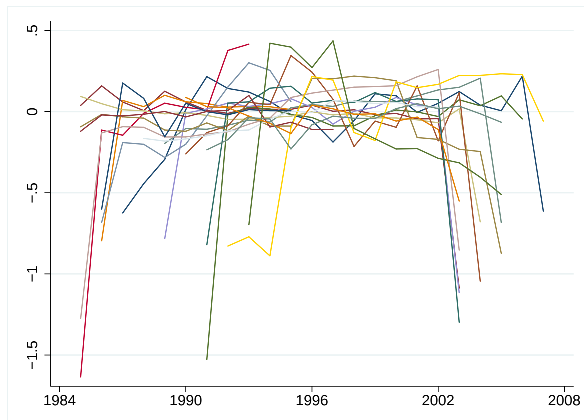
(a) Complete spells: start in 1984, end in 2008.



(b) Incomplete spells: start in 1984, end in 2001–2003.



(c) Incomplete spells: start in 1989–1991, end in 2008.



(d) Incomplete spells: start > 1984, end < 2008.

Figure 1: Randomly Selected Earnings Paths. German Data

at the start and/or the end of incomplete earnings spells are considerably lower on average and substantially more volatile than typical earnings observations. Our theoretical argument implies that the non-randomness of earnings surrounding missing observations in the unbalanced samples can induce the discrepancy between the estimates in levels and differences in the data from the unbalanced samples. The quantitative question is how large this effect is.

To provide an answer, we proceed in three steps. First, we quantify the contribution of the low mean and high variance of earnings surrounding missing observations in the unbalanced samples to the subset of theoretical autocovariance moments on which the identification argument in levels and differences is based, and confirm that they induce the observed discrepancy in the estimates. Second, using unbalanced samples, we drop a few observations at the start and end of the incomplete earnings histories, as well as observations surrounding missing records. We find that estimating the earnings process in levels and in differences on the remaining data yields virtually identical estimates of the variances of permanent and transitory shocks. Third, we simulate artificial data based on these estimates of the earnings process while replicating the structure of the unbalanced samples (by design of this experiment, first and last observations as well as those surrounding missing observations are not systematically different from observations in the rest of the earnings histories). We find no discrepancy of the estimates in levels and differences in these artificial data. We then draw an additional transitory shock ("rare transitory shock") at the start and end of the earnings history and surrounding missing observations to replicate the mean and the variance of earnings in those periods in the data. We find that in this case, the estimates of the variance of permanent and transitory shocks are very different when moments in levels and differences are used, but are very close to those in the data from the corresponding unbalanced samples.

The results of these experiments lead us to conclude that the discrepancy in the estimates of the earnings process (1) in growth rates and levels is indeed driven by its misspecification. The nature of the misspecification is surprisingly simple. It is driven by the high variance and the lower mean of the observations surrounding missing records. We show that an extended earnings process that includes these elements can be estimated in the data. Estimation of such an extended process results in similar parameters regardless of whether the moments in levels or differences are used.

In the second set of quantitative experiments in the paper, we consider survey data from the Panel Study of Income Dynamics. Survey data are generally inferior to administrative data for the narrow purpose of studying the earnings dynamics. However, it is indispensable for understanding the path-through of earnings shocks to consumption. Thus, our objective in this part of the paper is to assess whether the same mechanism described above is also responsible for the diverging estimates of variances of permanent/transitory shocks when targeting moments in growth rates and levels using survey data and the importance of accounting for it for understanding consumption responses to earnings shocks. To this end, we follow Blundell, Pistaferri, and Preston (2008) in estimating consumption insurance coefficients for permanent and transitory idiosyncratic earnings shocks measured by the fraction of those shocks that does not translate into movements in consumption. Using their male earnings data from the Panel Study of Income Dynamics (PSID), we find that the presence of rare transitory shocks at the start and end of earnings histories leads to a substantial upward bias in the estimated insurance against permanent shocks. We show theoretically that this bias is driven by the same forces that cause overestimation of the variance of permanent shocks using the earnings moments in growth rates. The rare transitory (and highly insurable) shocks are effectively "misinterpreted" by those moments as being permanent.

While the mechanism described in this paper is powerful in reconciling the estimates of the earnings process in growth rates and levels, it is not the only mechanism that can generate such discrepancy. For example, Hryshko and Manovskii (2016) show that this mechanism is quantitatively important but not sufficient to eliminate the full amount of discrepancy in the estimates of the stochastic process for household disposable income in PSID data. Instead, they argue, the remaining discrepancy is primarily driven by the typical restriction on the persistence of the permanent component of income, which limits its heterogeneity in the sample. Importantly, this type of misspecification cannot generate the difference between the theoretical moments that we use to establish identification in levels and differences in this paper, because they are identically affected by any such misspecification. These theoretical identifying moments can only differ if the underlying autocovariance moments on which they are based disagree, and we show that this is indeed the consequence of the low mean and high variance of observations at the start and end of earnings spells. We find that this accounts for virtually all discrepancy of the estimates in growth rates and levels in the earnings data we consider in this paper.

The rest of the paper is organized as follows. In Section 2, we discuss identification of the permanent-transitory decomposition of earnings, and derive theoretically the biases in the estimated variances of permanent and transitory shocks when using the moments in levels and differences constructed from an unbalanced panel. In Section 3, we describe the administrative Danish and German data and the estimation procedure. In the same section, we present basic estimation results and document that earnings are typically lower and more volatile in the periods surrounding missing observations. In Section 4, we show that this property of earnings quantitatively accounts for the difference in estimates of earnings processes in levels and differences in administrative data. In Section 5, we study theoretically and quantitatively the bias induced by this property of earnings on the estimated parameters of the earnings process when using survey data from the PSID and on the estimated insurance coefficients against permanent and transitory shocks. Section 6 concludes.

2 Sources of the Differences in Estimates Targeting Earnings Growth Rates and Levels

2.1 Identifying Moments

In the literature, estimation of the parameters of the earnings process typically relies on the minimum-distance method. In particular, estimation based on the moments in levels targets the entire set of autocovariance moments $E[y_{it}y_{it+j}]$, where $i \in [1, N]$ denotes individuals in the sample, t denotes time, and j denotes all leads and lags of earnings observed in the data. In differences, estimation targets the full set of autocovariance moments $E[\Delta y_{it} \Delta y_{it+j}]$, where Δ is the difference operator between two consecutive observations, so that $\Delta y_{it} \equiv y_{it} - y_{it-1}$ and $\Delta y_{it+j} \equiv y_{it+j} - y_{it+j-1}$.

Although all available autocovariance moments are used in estimation, identification is usually established using only a subset of autocovariance moments; see, e.g., Meghir and Pistaferri (2004), Blundell, Pistaferri, and Preston (2008), and Heathcote, Storesletten, and Violante (2014). For example, consider the earnings process that consists of a random walk and an i.i.d. transitory shock, which corresponds to setting $\theta(L)$ and ϕ_p to 1 in Eq. (1). This process was considered by Heathcote, Perri, and Violante (2010), who proposed the following moments to identify the variances of permanent and transitory shocks at time t:

`Differences:`

$$\sigma^2_{\xi,t} = E[\Delta y_{it}\Delta y_{it-1}] + E[\Delta y_{it}\Delta y_{it}] + E[\Delta y_{it}\Delta y_{it+1}], \tag{D1}$$

$$\sigma^2_{\epsilon,t} = -E[\Delta y_{it}\Delta y_{it+1}]. \tag{D2}$$

Note that (D1) and (D2) represent linear combinations of autocovariance moments for earnings growth rates. For clarity, we will refer to individual autocovariance moments as simply "moments," and to a linear combination of autocovariance moments used for identification such as (D1) and (D2) as "identifying moments."

Expanding (D1) and (D2), we obtain the identifying moments for the variances of permanent and transitory shocks, based on autocovariance moments in levels, at time t:

`Levels:`

$$\sigma^2_{\xi,t} = E[y_{it}y_{it+1}] - E[y_{it+1}y_{it-1}] - E[y_{it}y_{it-2}] + E[y_{it-1}y_{it-2}], \tag{L1}$$

$$\sigma^2_{\epsilon,t} = E[y_{it}y_{it}] - E[y_{it}y_{it+1}] - E[y_{it-1}y_{it}] + E[y_{it-1}y_{it+1}]. \tag{L2}$$

In a sample of individuals whose earnings are nonmissing for the periods $t-2$ through $t+1$, the identifying moments (D1)-(D2) and (L1)-(L2) are expected to deliver identical estimates of the variance of permanent and transitory shocks at time t, since they are based on exactly the same earnings information. Moreover, as the moments (L1)-(L2) simply represent an expansion of the moments (D1)-(D2), they will be identically affected by other potential misspecifications of the earnings process. This allows us to isolate and measure the importance of the high variance and low mean of the observations at the start and end of contiguous earnings spells, which, as we show below, contribute differently to the autocovariance moments on which (D1)-(D2) and (L1)-(L2) are based.[6]

[6]Identifying moments in levels can be constructed using fewer autocovariance moments, such as

$$\sigma^2_{\xi,t} = E[y_{it}y_{it+1}] - E[y_{it}y_{it-1}], \tag{L1-Short}$$

$$\sigma^2_{\epsilon,t} = E[y_{it}y_{it}] - E[y_{it}y_{it+1}]. \tag{L2-Short}$$

These moments are analogous to those in Heathcote, Perri, and Violante (2010) if one relies on the annual data, instead of biennial data used in their paper, for identification of the variances. These identifying moments in levels do not, however, use the same information as the identifying moments (D1)-(D2) in differences. For example, the information on earnings in $t-2$ is used in (D1) but not in (L1-Short). Moreover, Hryshko and

For example, the presence of omitted idiosyncratic trends in earnings will not induce a wedge between the estimated variances of permanent shocks using the moments (L1) and (D1) (or transitory shocks using the moments (L2) and (D2)). Specifically, suppose individuals differ in growth rates such that the earnings process is $y_{it} = \alpha_i + \beta_i h_{it} + p_{it} + \epsilon_{it}$, where $\beta_i \sim \text{iid}(0, \sigma_\beta^2)$ and h_{it} counts years of (potential) work experience. In this case (L1) and (D1) will both deliver $3\sigma_\beta^2 + \sigma_{\xi_t}^2$.[7] It follows that both (L1) and (D1) will recover an upward-biased estimate of the variance of the permanent shock but the bias will be the same in levels and differences. Relatedly, the typical estimates of σ_ξ^2 using (D1) imply a much steeper profile of earnings inequality over the life cycle (and time) than that observed in the data. The fit to this profile might be improved if one allows for a negative cross-sectional correlation between initial conditions, α_i, and permanent shocks, ξ_{it}. Omitting such correlation, however, does not induce a difference in the estimated moments (L1) and (D1). For example, suppose that the correlation is implied by $\xi_{it} = \kappa\alpha_i + \eta_{it}$, where η_{it} is orthogonal to α_i and ϵ_{it}. In this case, (D1) and (L1) will recover identical upward-biased estimate $3\kappa^2\sigma_\alpha^2 + \sigma_{\xi_t}^2$, but the bias will once again be the same in levels and differences.

Importantly, each autocovariance moment is measured as the average across all available observations that contribute to it. This implies that, although the identifying moments (D1)-(D2) and (L1)-(L2) are based on the same earnings data, the autocovariance moments used in estimating (D1)-(D2) and (L1)-(L2) are computed using different sets of observations. Returning to the extreme example used in the Introduction, consider an individual who appears in the sample only once, in period t. This individual will contribute to the autocovariance moment $E[y_{it}y_{it}]$, and thus his only earnings observation will affect the identifying moment (L2) but it will not contribute to any autocovariance moment used to construct the corresponding identifying moment in differences (D2). If earnings of individuals who appear in the sample only once are systematically different, this will induce the difference between identifying moments (L2) and (D2) and lead to different estimates of the variance of transitory shocks using the moments in levels and differences.

Similarly, we will now show that earnings observations at the time individuals enter or exit the sample contribute differently to the autocovariance moments on which the identifying moments (D1)-(D2) and (L1)-(L2) are based. Moreover, our empirical analysis will reveal that these earnings observations are systematically different (they are typically lower and substantially more volatile). In the rest of this section we formally show that this induces systematic differences in estimated variances of permanent and transitory shocks using the moments in growth rates and levels. In subsequent sections, we quantify the magnitude of the

Manovskii (2016) show that a misspecification of the persistence of the permanent component drives a wedge between the estimates based on identifying moments (D1)-(D2) and (L1-Short)-(L2-Short), but not between identifying moments (D1)-(D2) and (L1)-(L2).

[7] This derivation assumes that $\text{corr}(\alpha_i, \beta_i) = 0$ but a similar expression obtains if this assumption is relaxed.

induced difference.

2.2 The Effects of Rare Shocks in Various Samples

We will consider three types of samples. Consider a dataset with panel data on individual earnings that starts in period t_0 and ends in period T. We refer to the sample as *balanced* if all individuals in the sample have $T - t_0 + 1$ valid earnings observations. While not part of the formal definition, it is convenient to think that earnings spells of individuals in the balanced samples start before t_0 and end after T. In other words, the boundaries of the balanced sample mechanically truncate continuous earnings spells in progress. We refer to samples that include only uninterrupted earnings spells (i.e., no gaps) but with duration of less than $T - t_0 + 1$ for at least some individuals as consecutive unbalanced samples. Finally, we refer to unbalanced samples that also include individual earnings spells interrupted by missing observations in any period $t \in (t_0, T)$ as non-consecutive unbalanced samples.

2.2.1 Consecutive unbalanced samples

The nature of these samples is such that at least some individuals are observed starting or ending their earnings spells inside the sample window.

As mentioned above and documented below, earnings have a lower mean and are highly volatile in the first and last periods of an incomplete earnings history. Consider modeling this through an additional transitory shock that occurs only in the first and last year of an individual's earnings history, that is

$$y_{it} = \alpha_i + p_{it} + \epsilon_{it} + \nu_{it},$$

where ν_{it} has mean μ_ν (taking a negative value), and variance σ^2_ν and is uncorrelated with permanent and transitory shocks. Hereafter, we refer to the shock ν_{it} as a rare transitory shock, and call an earnings observation y_{it}, affected by this shock, an outlying earnings observation. We will now show that ignoring ν_{it} and estimating the process (1) instead leads to an upward bias in the estimated variance of permanent shocks using the moments in differences and in the estimated variance of transitory shocks using the moments in levels.

For simplicity, assume there is a set of individuals first entering the sample at time t, in the interior of the sample period $[t_0, T]$, whereas the remaining individuals are continuously observed throughout the sample. Individuals first appearing at time t will contribute to estimation of the autocovariance moments $E[y_{it}y_{it}]$ and $E[y_{it}y_{it+1}]$ in the identifying moment (L2). The estimated moment $E[y_{it}y_{it+1}]$ will be no different for such individuals than for the rest of the sample, and will equal $\sigma^2_\alpha + \text{var}(p_{it})$. The other moments in (L2), $E[y_{it-1}y_{it}]$ and $E[y_{it-1}y_{it+1}]$, will both equal $\sigma^2_\alpha + \text{var}(p_{it-1})$. The autocovariance moment $E[y_{it}y_{it}]$ estimated on the full sample, however, will equal $\sigma^2_\alpha + \text{var}(p_{it}) + \sigma^2_{\epsilon,t} + s_t(\mu^2_\nu + \sigma^2_\nu)$, where s_t is the share of individuals, at time t, whose (incomplete) spells start at time t in the total number of individuals at time t with nonmissing earnings. The identifying moment (L2), therefore, will recover an estimate of the variance of transitory shocks equal to $\sigma^2_{\epsilon,t} + s_t(\mu^2_\nu + \sigma^2_\nu)$, with an upward bias of $s_t(\mu^2_\nu + \sigma^2_\nu)$.

The variance of permanent shocks at time $t+1$, estimated using the identifying moment (D1), will also be biased upward. Individuals first appearing at t will contribute to estimation of the autocovariance moments $E[\Delta y_{it+1}\Delta y_{it+1}]$ and $E[\Delta y_{it+1}\Delta y_{it+2}]$ in the identifying moment (D1). For such individuals, the autocovariance moment $E[\Delta y_{it+1}\Delta y_{it+2}]$ will be no different from the rest of the sample and will equal $-\sigma^2_{\epsilon_{t+1}}$, while the autocovariance moment $E[\Delta y_{it+1}\Delta y_{it+1}]$ will equal $\sigma^2_{\xi_{t+1}} + s_{t,t+1}(\mu^2_\nu + \sigma^2_\nu) + \sigma^2_{\epsilon_t} + \sigma^2_{\epsilon_{t+1}}$, where $s_{t,t+1}$ is the share of individuals who start (incomplete) earnings spells at time t, with nonmissing earnings at times t and $t+1$, in the number of individuals with nonmissing earnings both at t and $t+1$. Since the autocovariance moment $E[\Delta y_{it+1}\Delta y_{it}]$ will be estimated using information for those individuals whose earnings are nonmissing in periods $t-1$ through $t+1$ and will equal $-\sigma^2_{\epsilon_t}$, the identifying moment (D1) for time $t+1$ will recover an estimate of the permanent shock equal to $\sigma^2_{\xi_{t+1}} + s_{t,t+1}(\mu^2_\nu + \sigma^2_\nu)$, with an upward bias of $s_{t,t+1}(\mu^2_\nu + \sigma^2_\nu)$.

Note that if the rare shock first appears, say, at time $t+1$, i.e. in the interior of an earnings spell for individuals first entering into the sample at time t, it will simply elevate, by the same magnitude, the estimated variance of transitory shocks in levels and differences at time $t+1$, with no differential effect on the identifying moments (L2) and (D1).

Summing up, incomplete earnings spells first appearing in the sample at t will bias upward the estimated variance of transitory shocks at time t when targeting the moments in levels, and will bias upward the estimated variance of permanent shocks at time $t+1$ when targeting the moments in differences. They have no effect, at any point in time, on the estimated magnitude of the identifying moments (L1) and (D2).

The same logic extends to the incomplete earnings spells ending at time t, which is different from the last potential sample year T – the presence of such spells will produce upward-biased estimates of permanent variances in differences at t (since these individuals will contribute to estimation of the moment $E[\Delta y_{it}\Delta y_{it}]$ that is part of the identifying moment D1) and of transitory variances in levels at t.

2.2.2 Non-consecutive unbalanced samples

We now consider the consequences of missing earnings in the interior points of the earnings history. We assume that individual earnings are realizations of the earnings process (1), with some observations missing in any period $t \in (t_0, T)$. We will show below that such periods are often associated in the data with low mean and high variance of earnings in periods $t-1$

and $t + 1$. We model this by introducing additional rare transitory shocks with a negative mean μ_ν at the time before and after earnings are missing (ν_{it-1} and ν_{it+1}, respectively) that are assumed to be uncorrelated with permanent and transitory shocks, and uncorrelated with each other:[8]

$$
\begin{aligned}
&y_{it-1} = \alpha_i + p_{it-1} + \epsilon_{it-1} + \nu_{it-1}, \\
&y_{it} \ \text{ missing}, \\
&y_{it+1} = \alpha_i + p_{it+1} + \epsilon_{it+1} + \nu_{it+1}.
\end{aligned}
$$

Assume there is a set of individuals whose earnings are missing at time t, which is interior to the sample period $[t_0, T]$, while the rest of individuals have continuously observed earnings throughout the whole sample period.

In this case, the variance of transitory shocks at times $t-1$ and $t+1$ using the moments in levels will be biased upward as the autocovariance moments $E[y_{it-1}y_{it-1}]$ and $E[y_{it+1}y_{it+1}]$ in the identifying moment (L2) are amplified by the variation of the rare shocks. Similarly, the variance of permanent shocks at times $t-1$ and $t+2$ using the moments in differences will be biased upward as the autocovariance moments $E[\Delta y_{it-1}\Delta y_{it-1}]$ and $E[\Delta y_{it+2}\Delta y_{it+2}]$ in the identifying moment (D1) are amplified by the variation of the rare shocks. Since the rare shocks are assumed to be uncorrelated, the identifying moments (L1) and (D2) will not be affected.

Thus, incomplete earnings spells with missing earnings at t, in the interior of the sample period, will bias upward the estimated variance of transitory shocks at times $t-1$ and $t+1$ when targeting the moments in levels, and will bias upward the variance of permanent shocks at times $t-1$ and $t+2$ when targeting the moments in differences.

2.3 Extensions

2.3.1 Limited persistence of ξ_{it} shocks

If ϕ_p in Eq. (1) is less than 1, one must rely on a modified set of identifying moments to recover the permanent and transitory variances. For a given estimate of the persistence ϕ_p,

[8] For ease of exposition, we assume that the mean and variance of the rare shock one year before and after earnings are missing are the same, although in the data they slightly differ.

which can be separately identified,[9] the set of identifying moments will amount to

`Differences:`

$$\sigma^2_{\xi,t} = E[\tilde{\Delta} y_{it}\tilde{\Delta} y_{it+1}] + \phi_p E[\tilde{\Delta} y_{it}\tilde{\Delta} y_{it}] + \phi_p^2 E[\tilde{\Delta} y_{it}\tilde{\Delta} y_{it-1}], \qquad \text{(D1-a)}$$

$$\sigma^2_{\epsilon,t} = -\frac{1}{\phi_p} E[\tilde{\Delta} y_{it}\tilde{\Delta} y_{it+1}], \qquad \text{(D2-a)}$$

where $\tilde{\Delta} y_{it} \equiv y_{it} - \phi_p y_{it-1}$.

Expanding the above moments results in the following set of moments in levels identifying the variances at time t:

`Levels:`

$$\sigma^2_{\xi,t} = E[y_{it}y_{it+1}] - \phi_p E[y_{it+1}y_{it-1}] - \phi_p^3 E[y_{it}y_{it-2}] + \phi_p^4 E[y_{it-1}y_{it-2}], \qquad \text{(L1-a)}$$

$$\sigma^2_{\epsilon,t} = E[y_{it}y_{it}] - \frac{1}{\phi_p} E[y_{it}y_{it+1}] - \phi_p E[y_{it-1}y_{it}] + E[y_{it-1}y_{it+1}]. \qquad \text{(L2-a)}$$

Although the biases for the variance of transitory shocks in levels will be exactly the same as in the random-walk case, the biases for the variance of permanent shocks recovered using the identifying moments in differences will be scaled by the persistence ϕ_p. Note, however, that the bias will remain large, since ϕ_p is typically estimated at high values in various datasets.

2.3.2 Serially correlated transitory component and/or rare shocks

The transitory component is often estimated to have some persistence. Assume that the transitory component is modeled as $\tau_{it+1} = \epsilon_{it+1} + \theta_\tau \epsilon_{it}$, and that the rare-shock component is modeled as $\chi_{it} = \nu_{it}$, which is nonzero in the beginning and/or end of an incomplete earnings spell, and before/after a missing earnings record, and that $\chi_{it+1} = \theta_\chi \nu_{it}$ – both will be consistent with the autocovariance function for earnings growth rates truncating at the second order, as is often found in the empirical applications.[10] In this case, the moments (L1)–(D2) no longer identify the variances of permanent and transitory shocks. In growth rates, the identifying moment for the variance of permanent shocks should be modified to

$$\sigma^2_{\xi,t} = E[\Delta y_{it}\Delta y_{it+2}] + E[\Delta y_{it}\Delta y_{it+1}] + E[\Delta y_{it}\Delta y_{it}] + E[\Delta y_{it}\Delta y_{it-1}] + E[\Delta y_{it}\Delta y_{it-2}]. \qquad \text{(D1-b)}$$

The variance of permanent shocks at time $t+1$, estimated using (D1-b), will be biased upward by the magnitude $s_{t,t+1}(1-\theta_\chi)^2(\mu^2_\nu + \sigma^2_\nu)$ for a sample with consecutive earnings spells where

[9] The persistence ϕ_p can be recovered from the moments $\frac{E[y_{it+k+3}y_{it+k}] - E[y_{it+k+2}y_{it+k}]}{E[y_{it+k+2}y_{it+k}] - E[y_{it+k+1}y_{it+k}]}$ for $k \geq 0$. One can also use the moments in growth rates to identify it; see, e.g., Hryshko (2012). There is also a large literature, reviewed in MaCurdy (2007) and Arellano and Honoré (2001), that does not rely on fitting the autocovariance function of earnings but exploits various orthogonality conditions in a GMM setting to recover the persistence.

[10] This formulation assumes that ν- and ϵ-shocks both die out in two periods, with the difference that the rare-shock process does not renew itself in the next period with a new ν-shock.

a fraction of individuals enter the sample at time $t > t_0$, for the first time. Note that the bias will remain large for small positive values of θ_χ. If, instead, individuals exit the sample at some time $t < T$, the bias of the permanent variance using the moments in growth rates will be unaffected by serial correlation of the rare shocks since the earnings paths for such individuals are unobserved past year t; the bias in this case will be the same as in the case of a serially uncorrelated transitory component. The same logic extends to the biases in the non-consecutive samples. The variance of permanent shocks recovered using the moments in levels will remain unbiased (as can be verified from the identifying moment for permanent shocks in levels obtained by expanding (D1-b)).

Under assumption of no measurement error in administrative earnings, θ_τ can be identified from the first and second-order autocovariances in earnings growth rates if the transitory component is serially correlated and there are no rare shocks; see, e.g., Meghir and Pistaferri (2004). One can then identify the variance of transitory shocks dividing (L2) and (D2) by $(1-\theta_\tau)^2$. If the rare shock is serially correlated, however, θ_τ will be recovered with a bias using the standard moment. We will label this estimate as $\tilde{\theta}_\tau$. Assuming that the variance of transitory shocks does not change much between adjacent periods, for the data with incomplete consecutive spells that start at t, an estimate of the variance of transitory shocks relying on (L2) will yield $(1-\tilde{\theta}_\tau)^{-2}\left[(1-\theta_\tau)^2\sigma^2_{\epsilon t} + s(1-\theta_\chi)(\mu^2_\nu + \sigma^2_\nu)\right]$, whereas an estimate relying on (D2) will yield an estimate $(1-\tilde{\theta}_\tau)^{-2}\left[(1-\theta_\tau)^2\sigma^2_{\epsilon_t} - s\theta_\chi(1-\theta_\chi)(\mu^2_\nu + \sigma^2_\nu)\right]$ for $t+1$.[11] Clearly, an estimate of the variance of transitory shocks in levels is larger than an estimate using growth rates given θ_χ is nonnegative. This logic extends to other examples of incomplete earnings spells in consecutive and non-consecutive panels – the estimated variance of transitory shocks using the moments in levels will be higher than the estimated variance of transitory shocks using the moments in growth rates.

2.4 Summary

The analysis above yields three major implications if rare shocks are present in the data. First, estimating the abbreviated earnings process in (1), one may expect to recover without any biases the variance of transitory shocks using the moments in growth rates if the rare shock is not serially correlated, and the variance of permanent shocks using the moments in levels. Second, the identifying moments in levels tend to produce upward-biased estimates of the variance of transitory shocks, while the identifying moments in differences produce upward-biased estimates of the variance of permanent shocks. The magnitude of the biases depends positively on the variance of the rare shocks and on the difference between their mean from the mean of the shocks in the rest of earnings histories. Finally, if one's interest extends beyond identifying properties of permanent and transitory shocks of the abbreviated earnings

[11] We assumed that $s_t = s_{t,t+1} = s_{t,t+2} = s$ in the derivation.

process in (1), the remaining parameters of the comprehensive earnings process can also be estimated by introducing the moments identifying the mean and variance of rare shocks.

3 Data, Estimation Details, and Basic Results

3.1 Data

In this section we describe the administrative data and construction of the samples that we study. Following the literature, we focus on individuals with a strong attachment to the labor market characterized by sufficiently high earnings and time spent working.[12]

3.1.1 Danish data

Several administrative registers provided by Statistics Denmark were used to construct our samples. The tax register from 1980–2006 provides panel data on total earnings for more than 99.9 percent of Danish residents between the ages of 15 and 70. The register was merged with the Danish Integrated Database for Labor Market Research (IDA) so that additional demographic variables such as educational status could be appended. The population consists of Danish males born in 1951 through 1955. We observe annual earnings over the period of 1980 through 2006. We first remove all individuals who were ever self-employed and drop records in which an individual was making non-positive labor market earnings. Next, we drop records for those individuals who have worked less than 10 percent of the year as a full-time employee; this restriction limits our data to the period 1981–2006, since we cannot identify full-time employment status for the year 1980.[13] Annual earnings in a particular year

[12]The selection rules we adopt are typical of the literature that utilizes survey data as well as administrative data. For example, Guvenen, Ozcan, and Song (2014) use U.S. administrative data on individual wage and salary income and make the following sample selection: "For a statistic computed using data for not necessarily consecutive years $t_1, t_2, \ldots, t_n$, an individual observation is included if the following three conditions are satisfied for all these years: the individual (i) is between the ages of 25 and 60, (ii) has annual wage/salary earnings that exceed a time-varying minimum threshold, and (iii) is not self-employed (i.e., has self-employment earnings less than the same minimum threshold). This minimum, denoted $Y_{min,t}$, is equal to one-half of the legal minimum wage times 520 hours... This condition allows us to focus on workers with a reasonably strong labor market attachment and avoids issues with taking the logarithm of small numbers. It also makes our results more comparable to the income dynamics literature, where this condition is standard." Similarly, DeBacker, Heim, Panousi, Ramnath, and Vidangos (2013) "...exclude earnings (or income) observations below a minimum threshold..." and "...take the relevant threshold to be one-fourth of a full-year, full-time minimum wage." In line with our selection of consecutive unbalanced samples (with the difference that we use at least nine consecutive earnings observations), Blundell, Graber, and Mogstad (2015) "...restrict the sample to individuals with at least four subsequent observations with positive market income."

[13]We use the variable "erhverv" from the IDAP table provided by Statistics Denmark. This variable calculates work experience as a full-time employee since 1980 based on individuals' yearly pension contributions and is available for all members of the population (with the exception of those individuals who have spent time abroad, for whom the variable is reset to 0). By taking the first difference of this measure, we can calculate the percentage of the year during which an individual has worked full-time, which restricts our observation period to 1981–2006.

include all earned labor income, taken from tax records, for that calendar year. This variable is considered "high quality" by Statistics Denmark in that it very accurately captures the earnings of individuals. Earnings are expressed in 1981 monetary units (Danish kroner). We calculate the maximum number of consecutive periods in which an individual has nonmissing earnings and use this information to construct two consecutive samples: a sample in which an individual's maximum spell is at least nine consecutive periods (102,825 individuals), and a balanced sample in which the individual's maximum spell covers the entire 26 periods (67,008 individuals). For the sample with nine or more consecutive observations, periods outside of the longest spell are dropped. Within the longest spell, an earnings outlier is defined by an increase in earnings of more than 500 percent or a fall of more than 80 percent in adjacent years. Individuals with earnings outliers within their longest spell are dropped. The third sample we consider consists of individuals who have at least 20 not necessarily consecutive periods in which they have nonmissing earnings (90,668 individuals). We also drop individuals from this sample if they have earnings growth outliers. Finally, we drop individuals if their educational status has changed during the spells considered. Table A-1 contains basic statistics for selected samples.

3.1.2 German data

We use administrative data from the IABS, a 2% random sample of German social security records for the years 1974–2008. A detailed description of the dataset can be found in Dustmann, Ludsteck, and Schönberg (2009). We use full-time job spells for German males born in 1951–1955, dropping the spells in East Germany. We also drop annual records when an individual was in apprenticeship during any part of the year. Individual real earnings are the sum of earnings from all jobs held within a year expressed in 2005 euros. We set individual education to the maximum schooling attained during the sample years, and set the number of days worked to the sum of calendar days on all jobs within a year. As individual earnings are right-censored at the highest level subject to social security contributions, we impute earnings exceeding the limit assuming that daily wages in the upper tail follow a Pareto distribution, the parameters of which differ by year and age group.[14] After 1983, earnings include one-time payments such as bonuses. To make variable definitions consistent throughout, we use only the data since 1984. We also drop individual records on annual earnings if the combined

[14]We consider the following eight age groups: those younger than 25, six five-year age groups (25–29, 30–34, 35–39, 40–44, 45–49, and 50–54), and those older than 54. We use a "fixed effects" imputation, keeping a uniform draw for each individual affected by the right-censoring limit fixed when creating a Pareto variate in different years. We also experimented with imputation based on the assumption that truncated log-wage distribution is normal, and a simpler imputation when daily wage is multiplied by the factor 1.2 if it hits the upper censoring limit. These three imputation methods have been used in Dustmann, Ludsteck, and Schönberg (2009). Our conclusions below are robust with respect to the choice of the imputation method as well as with respect to limiting the sample to individuals whose earnings histories are not affected by the censoring.

duration of job spells within a year is fewer than 35 calendar days, and drop records with very low daily earnings.[15] As in the Danish data, we construct three samples – balanced, with nine or more consecutive, and with 20 or more not necessarily consecutive earnings observations – and, as with the Danish samples, drop individuals who have earnings growth outliers. The respective samples contain 9,452, 18,130, and 13,635 individuals with 236,300, 379,080, and 330,748 observations, respectively. Table A-2 provides some descriptive details of the samples.

3.2 Estimation Details

As is standard in the literature, we estimate the earnings process in Eq. (1) using the method of minimum distance, fitting the data autocovariance function of log-earnings in levels or first differences to the autocovariance function implied by the model.[16] We allow for an MA(1) transitory component and an unrestricted estimation of the persistence of the permanent component, ϕ_p.[17] Thus, we estimate five parameters in total – the persistence and the variance of permanent shocks, ϕ_p and σ^2_ξ; the persistence and the variance of transitory shocks, θ and σ^2_ϵ; and the variance of individual fixed effects, σ^2_α. We assume that individuals start accumulating permanent and transitory shocks at the age of 25 so that part of the estimated variance of fixed effects captures the accumulated permanent and transitory components prior to that age. We remove predictable variation in earnings by estimating cross-sectional regressions of log earnings on educational dummies, a third polynomial in age, and the interactions of the age polynomial with the educational dummies. Our measure of idiosyncratic earnings, consistent with the literature, is the residual from those regressions. Since our samples are large, we estimate the model using the optimal weighting matrix which is an inverse of the variance-covariance matrix of the data moments.

3.3 Basic Results

3.3.1 Samples with nine or more consecutive observations

Columns (1)–(4) in Table 1 contain estimation results for the samples with nine or more consecutive observations in the German and Danish data.[18] The permanent component is estimated to be close to a random walk using the moments in differences, but slightly less persistent using the moments in levels. Importantly, in both datasets the variance of the permanent shock is about two times larger in the estimation that uses the moments in growth

[15] The highest marginal part-time income threshold during the sample period was 13.15 euros a day (set for the first time in 2003), and we drop the records with daily earnings below 14 euros in 2003 prices in any year.

[16] One of the recent exceptions is Browning, Ejrnæs, and Alvarez (2010) who, apart from selected moments in levels and differences, fit a variety of other data moments studied in the literature on earnings dynamics.

[17] In the previous version of the paper, we allowed for an AR(1) transitory component instead with little influence on the results.

[18] In differences, the variance of fixed effects is not identified.

rates, while the estimated variance of the transitory shock is larger using the moments in levels. Thus, our administrative data exhibit the same large discrepancy that is endemic in this literature. The pattern is less pronounced in the Danish data which is consistent with the mechanism we describe. In the Danish data, 65% of individuals have complete earnings spells, while in the German data this number is only 52%. Consequently, fewer individuals have outlying earnings observations adjacent to missing ones in the Danish data.[19]

3.3.2 Samples with 20 or more not necessarily consecutive observations

Columns (5)–(8) in Table 1 contain the results for the samples with 20 or more not necessarily consecutive observations. The variances of persistent shocks are somewhat smaller than those in columns (1)–(4), whereas the variances of transitory shocks are similar in magnitude. Importantly, we still observe that estimations using the moments in differences deliver relatively higher estimates of the variance of permanent shocks, while estimations in levels deliver relatively higher estimates of the variance of transitory shocks, once again confirming the widely documented discrepancy.[20]

3.3.3 Balanced samples

Estimation results based on the balanced samples are reported in columns (9)–(12) of Table 1. Relative to the estimates on the unbalanced samples discussed above, the use of balanced samples results in a more than 50% reduction of the variance of permanent shocks when using the moments in differences. There is a similarly striking reduction of at least 50% in the variance of transitory shocks when using the moments in levels. It appears that the use of balanced samples largely eliminates the discrepancy between the estimates of the earnings process in levels and differences.

3.4 A Closer Look at Unbalanced Samples

The results of estimation on balanced and unbalanced samples indicate that the discrepancy between the estimates based on the moments in levels and differences is specific to unbalanced samples. One possible explanation for this finding is that individuals with shorter earnings spells are intrinsically different, and that while permanent/transitory decomposition in Eq. (1) is appropriate for workers in the balanced sample, it provides a fundamentally misspecified model of the earnings processes for individuals in the unbalanced samples. Alternatively,

[19] Randomly dropping individuals with incomplete earnings histories in the German data to match their share in the Danish data results in similar discrepancies across the two datasets.

[20] As was the case with nine or more consecutive observations, the discrepancy is less pronounced in the Danish data because the share of individuals with complete earnings spells is larger, and the share of missing earnings observations in the potential number of earnings observations (calculated as $(T - t_0 + 1) \times N$, where N is the number of individuals in a sample) is smaller than in the German data.

it is possible that the decomposition is essentially valid but that individuals in unbalanced panels either have higher shock variances or represent a selection of workers who experienced earnings "shocks" unfavorable enough to push them out of employment. One consequence of such selection is that the earnings surrounding the missing observations are likely to belong to workers in transit into or out of employment, with a potentially large impact on earnings in those periods. As discussed in Section 2, this can induce the difference in the estimates of the earnings process in growth rates or levels.

Indeed, Figure 1 in the Introduction revealed a clear pattern that the earnings at the start and/or the end of incomplete earnings spells are considerably lower on average and substantially more volatile than typical earnings observations. To explore these patterns more formally, in columns (1)–(4) of Table 2 we report the estimates from the fixed-effects panel regressions of residual earnings on dummies for the first and last years of individual earnings spells inside the overall sample window. Specifically, the dummies "Year observed: first"−"Year observed: third" equal one if an individual's first earnings record in the sample occurs later than 1984 in the German data (later than 1981 in the Danish data), and zero otherwise, while the dummies "Year observed: second-to-last"−"Year observed: last" equal one if an individual's last earnings record is prior to 2008 in the German data (2006 in the Danish data), and zero otherwise.[21]

In both samples and both datasets, earnings are about 0.50 to 0.60 log points lower than an individual's average in the first year of the spell, whereas the last earnings record is below an individual's average by about 0.30 to 0.40 log points. Earnings are still lower in the two years following the first earnings record as well as in the two years preceding the last earnings record. Moreover, earnings are, on average, also lower in the years preceding and following a missing earnings record in the non-consecutive samples. Clearly, the "shock" in the first year of an individual's spell is transitory, but somewhat persistent.[22] Interestingly, the dummies for the few first and last earnings records within a spell explain 5 to 13 percent of the variation in residual earnings. This number is quite high taking into account that a variety of observable factors normally explain about 30 percent of variation in earnings.

Performing the same experiment in reverse, we use our samples with 20 or more not necessarily consecutive observations to assess the predictive power of earnings dynamics for the incidence of missing earnings. Specifically, in Table A-4, the dependent variable is a dummy that equals 100 if individual earnings are missing and 0 otherwise. We find that the predic-

[21] To reinforce the conclusion that patterns in Table 2 are actually driven by starting and ending of the earnings spells, in Table A-3 we repeat the same analysis by focusing on individuals whose earnings spells begin in the first sample year or end in the last sample year. For the vast majority of these individuals, such cutoffs do not represent an actual start or end of their earnings spells; instead, the sample window mechanically truncates earnings spells in progress. Accordingly, the first (last) few dummies equal one if an individual's fist (last) earnings record is in the first (last) sample year, and zero otherwise.

[22] If the shock were permanent, it would elevate earnings in all periods, with no distinguishable differences in the first earnings record from the individual's average.

tive power of observables – earnings growth rates before and after missing earnings records, together with education dummies and age – on the incidence of missing earnings is quite small, in line with Fitzgerald, Gottschalk, and Moffitt (1998) who made a similar observation using PSID data. The strong (weak) earnings growth after (before) missing records lacks high explanatory power for a missing record because there are also many declines and subsequent recoveries of earnings inside uninterrupted earnings spells, as can be seen, e.g., in Figure 1 panel (a). Nonetheless, missing observations are associated with positive earnings growth in the periods following a missing record and with negative earnings growth in the periods preceding a missing earnings record, implying that these individual realizations of residual earnings are not random draws from the earnings distribution. As pointed out by Moffitt and Gottschalk (2012), little is known about the effect of attrition on the autocovariance function of earnings and, therefore, on the estimates of the earnings process. Our results indicate that the effect can be large.

In columns (5)–(8) of Table 2, we proceed to explore the volatility of idiosyncratic earnings at the start and end of earnings spells. The size of squared residual earnings is mechanically higher in the few first and last earnings records since, as we have just seen, residual earnings are more negative, on average, in those periods. To remove the influence of more negative residual earnings in those periods, we take the (individually demeaned) residuals from the regressions of columns (1)–(4), and then square them. In the German data, the overall mean of squared residual earnings is about 0.15 in both samples while in the Danish data, the corresponding mean in both samples is 0.11. The results imply that earnings are significantly more volatile in the (few) first and last years of individual spells. For example, in the German data, the mean of squared residual earnings in the first year is about 153% ($100 \times 0.23/0.15$) larger than the typical size measured by the mean of squared residual earnings in the sample. In the German consecutive sample, about 23% of individuals have their first earnings record after 1984, the first calendar year of the sample, and about 31% of individuals have their last record before 2008, the last year of the sample. The same numbers for Danish data are 18% and 22%, respectively. This is a non-trivial number of individuals with pronounced differences in the level and volatility of residual earnings in the few first and last periods of earnings spells. In the non-consecutive samples, earnings in the periods preceding and following interior missing earnings records are also highly volatile. In the German data, for instance, the volatility of earnings observations one year before a missing record is about 100% ($100 \times 0.15/0.15$) larger than the volatility of typical earnings observations. Tables A-1 and A-2 indicate that the fraction of missing earnings in the non-consecutive samples is also quite large – over 5% in the German data and 14% in the Danish data.

Finally, we consider some of the economic forces leading to low and volatile earnings at the start and end of the earnings spells. One obvious explanation is based on the fact that the data on earnings are typically recorded at an annual frequency. An individual who is, say, entering the sample for the first time is (statistically) expected to enter in the middle of the year, but may enter at any point throughout the year. Thus, earnings in that year are expected to be lower and have a larger variance than interior earnings observations from contiguous earnings histories. We can assess this conjecture using our German data which contain information on the number of days worked on all jobs and the average daily wage from all jobs held during a year. We use these data to decompose earnings cuts in the years around missing earnings records due to reduction in days worked and wages. As can be seen from Table A-5, most of the reduction in earnings in the first or last year of the earnings spell is due to the reduction in days worked. The reduction in wages in those years is non-trivial as well, but becomes even more relatively important for earnings fluctuations two or three years away from the start or the end of the earnings spell.[23]

In Table A-6, we report that years at the start and at the end of earnings spells are associated with a significant increase in the probability of occupation (three-digit) and industry (two-digit) change as well as a higher incidence of unemployment. On average, the fraction of individuals, in the cohort we study, who change occupation in a given year is 5%, declining over the life cycle from 8% among those observed in 1984 (at ages 29–33) to 2% in 2008 (at ages 53–57). For those individuals who enter the sample later than 1984, the probability of an occupational switch increases (controlling for age effects) by about 31 percentage points in the first year. A qualitatively similar picture holds for changes in industrial affiliation (on average, individuals change industry about 4% of the time). Individuals continue experiencing substantial occupation and industry mobility in the second and third years of incomplete earnings spells, as well as in a few years prior to the end of the incomplete earnings spell. This is consistent with a stronger labor-market attachment in the second and third years (and next-to-last years) of the incomplete spells when the reduction in days worked is not as large as the reduction in daily wages. In columns (3) and (6) of Table A-6, we show that individuals are also substantially more likely to have spent part of the year unemployed in the few years surrounding missing earnings records (on average, individuals in our sample are unemployed about 3% of the time). In summary, outlying earnings records around the missing ones are due to variation both in hours and wages associated with job, occupation, and industry mobility, and lost work time due to spells of unemployment.

[23]Consistently with this interpretation, we find much smaller differences in the estimated variances of permanent and transitory shocks when using the moments in levels and differences in *wages*, as much of the variability in earnings in our administrative datasets at the start and end of contiguous spells is due to the variability in hours. This finding is consistent with the observation of Krueger, Perri, Pistaferri, and Violante (2010) that the discrepancy is larger for the estimates of earnings processes than wage processes in a broad cross-section of countries.

4 Quantitative Evaluation of the Mechanism

4.1 Direct Evaluation of the Biases Using the Permanent-Transitory Decomposition Moments

In this section, we directly verify that outlying observations induce most of the difference between permanent and transitory shock variances implied by identifying moments (L1)–(L2) and (D1)–(D2). As an example of computing these implied variances, we calculate an estimate of the permanent variance at time t using the identifying moment in levels (L1) as

$$\sigma^2_{\xi,l,t} = \frac{\sum_i y_{i,t}y_{i,t+1}}{\sum_i I^i_{t,t+1}} + \frac{\sum_i y_{i,t-2}y_{i,t-1}}{\sum_i I^i_{t-2,t-1}} - \frac{\sum_i y_{i,t+1}y_{i,t-1}}{\sum_i I^i_{t-1,t+1}} - \frac{\sum_i y_{i,t}y_{i,t-2}}{\sum_i I^i_{t-2,t}}, \tag{2}$$

where the subscript l indicates that we are estimating the variance using information on log-earnings in levels, and $I^i_{t,t'}$ is an indicator function taking the value of one if individual earnings observations are nonmissing in both years t and t', and taking the value of zero otherwise. Note that individual i will not contribute to the estimated variance of the permanent shock at time t only if all of the earnings cross-products for that individual – $y_{it}y_{it+1}$, $y_{it-2}y_{it-1}$, $y_{it+1}y_{it-1}$, and $y_{it}y_{it-2}$ – are missing.

Let I^m_{it} be an indicator function that equals one if an individual's earnings is missing at one of the periods $t-2,\dots,t+2$ defined as $\mathbb{1}\left(\sum_{j=-2}^{2}(1-I_{it+j})>0\right)$, where $\mathbb{1}(\cdot)$ is an indicator function that equals one if the expression in brackets is true and zero otherwise. We calculate the variance of permanent shocks due to outlying observations surrounding the missing earnings records, $\sigma^2_{\xi,l,o,t}$, as

$$\sigma^2_{\xi,l,o,t} = \frac{\sum_i y_{i,t}y_{i,t+1}I^m_{it}}{\sum_i I^i_{t,t+1}I^m_{it}} + \frac{\sum_i y_{i,t-2}y_{i,t-1}I^m_{it}}{\sum_i I^i_{t-2,t-1}I^m_{it}} - \frac{\sum_i y_{i,t+1}y_{i,t-1}I^m_{it}}{\sum_i I^i_{t-1,t+1}I^m_{it}} - \frac{\sum_i y_{i,t}y_{i,t-2}I^m_{it}}{\sum_i I^i_{t-2,t}I^m_{it}}. \tag{3}$$

An estimate of the permanent variance in levels, net of the effects of outliers, $\sigma^2_{\xi,l,n,t}$, can then be calculated as

$$\begin{aligned}\sigma^2_{\xi,l,n,t} = &\frac{\sum_i y_{i,t}y_{i,t+1}(1-I^m_{it})}{\sum_i I^i_{t,t+1}(1-I^m_{it})} + \frac{\sum_i y_{i,t-2}y_{i,t-1}(1-I^m_{it})}{\sum_i I^i_{t-2,t-1}(1-I^m_{it})} \\ &- \frac{\sum_i y_{i,t+1}y_{i,t-1}(1-I^m_{it})}{\sum_i I^i_{t-1,t+1}(1-I^m_{it})} - \frac{\sum_i y_{i,t}y_{i,t-2}(1-I^m_{it})}{\sum_i I^i_{t-2,t}(1-I^m_{it})}.\end{aligned} \tag{4}$$

We can similarly define the variances of permanent and transitory shocks in levels and differences for the consecutive unbalanced panels – e.g., the permanent variance utilizing all sample information ($\sigma^2_{\xi,l,t}$ for levels and $\sigma^2_{\xi,d,t}$ for differences), the permanent variance due to outlying observations in the first and last few periods of an individual's earnings spell ($\sigma^2_{\xi,l,o,t}$ and $\sigma^2_{\xi,d,o,t}$), and the permanent variance net of outlying effects ($\sigma^2_{\xi,l,n,t}$ and $\sigma^2_{\xi,d,n,t}$).

We present the estimates of those variances, averaged across all sample years, for both datasets in Table 3. For the German data, in the consecutive sample, the estimates of the variance of permanent shocks in levels and differences using all sample information are 0.013 and 0.024, respectively.[24] Net of outliers, the estimated variances are $\hat{\sigma}^2_{\xi,l,n} = 0.010$ in levels and $\hat{\sigma}^2_{\xi,d,n} = 0.010$ in differences. The unadjusted variances of transitory shocks in levels and differences are estimated at 0.020 and 0.008, respectively, while the variances net of outliers in levels and differences are both estimated at 0.007. The results for the Danish data are qualitatively similar. Clearly, the discrepancy between the estimates of permanent and transitory shock variances in levels and differences is virtually eliminated when netting out the effects of outlying observations.

Similarly, in the German non-consecutive sample, the variances of permanent shocks are $\sigma^2_{\xi,l} = 0.0096$, $\sigma^2_{\xi,l,n} = 0.0097$, $\sigma^2_{\xi,d} = 0.018$, $\sigma^2_{\xi,d,n} = 0.0097$, while the variances of transitory shocks are $\sigma^2_{\epsilon,l} = 0.018$, $\sigma^2_{\epsilon,l,n} = 0.007$, $\sigma^2_{\epsilon,d} = 0.007$, $\sigma^2_{\epsilon,d,n} = 0.007$. Netting out the influence of missing observations and the influence of the first and last records in the earnings spells eliminates most of the discrepancy between the variances of permanent and transitory shocks in differences and levels.

4.2 Restricting Unbalanced Samples

One approach to eliminating the impact of low mean and high variance of observations at the start and end of earnings spells on the estimates of the permanent/transitory decomposition is to simply drop those observations. Accordingly, in Tables 4 and 5, columns (3) and (4) of Panel A, we repeat our analysis of Table 1 using the German and Danish samples with nine or more consecutive observations after dropping the first three observations for individuals whose earnings spells start after the first sample year and the last three observations for individuals whose earnings spells end before the last sample year.[25] In the same columns in Panel B we, in addition, drop three observations before and after a missing earnings record in the non-consecutive samples.

For the sample with nine or more consecutive observations, doing so barely affects the persistence of permanent shocks, although their variance estimated using the moments in differences is reduced by about 70%. The variance of transitory shocks estimated using the moments

[24]The estimates deviate from the values in Table 1 because we do not impose the exact permanent-transitory decomposition on the data in the minimum-distance estimation of Table 1. The difference between the estimated variance of permanent shocks in levels and differences is not as drastic as in Table 1 because the estimated persistence of the permanent shocks in levels is estimated to be lower than in differences in the minimum-distance estimation.

[25]Geweke and Keane (2000) drop the first earnings record in their analysis using the PSID because individuals are likely to work only part of the year the first time they are observed which is consistent with our results. Dropping the first record, however, may not be enough to eliminate the biases as the records in a few subsequent years may be different from the rest of earnings observations, reproducing the bias that had been created by the (dropped) first earnings record.

in levels is reduced by about 60%. We observe a similar pattern in the non-consecutive sample in Panel B. As a result, the estimated earnings process is virtually identical in estimations utilizing the moments for growth rates and levels.

A comparison with Table 1 also indicates, consistent with the analysis in Section 2, that in both datasets the variance of the permanent component is more robustly estimated using the moments in levels, whereas the variance of the transitory component, net of the transitory variation in earnings due to rare shocks, is more robustly estimated using the moments in differences.

4.3 Modeling Outlying Earnings Records

In columns (5)–(8) of Tables 4 and 5, instead of dropping outlying observations, we estimate an extended earnings process that explicitly models them. Specifically, we estimate the following model:

$$\begin{aligned}
y_{it} &= \alpha_i + p_{it} + \tau_{it} + \chi_{it}, \quad t = t_0, \dots, T \\
p_{it} &= \phi_p p_{it-1} + \xi_{it} \\
\tau_{it} &= \epsilon_{it} + \theta \epsilon_{it-1} \\
\chi_{it+j} &= \begin{cases} \nu_{it} & \text{if } y_{it-k} \text{ or } y_{it+k} \text{ is missing and } t-k \geq t_0,\ t+k \leq T, j = 0 \\ \theta \nu_{it} & j = 1 \\ 0 & \text{otherwise.} \end{cases}
\end{aligned} \tag{5}$$

In columns (5) and (6) of Tables 4 and 5, $k = 1$, i.e., we isolate only the first and last observation of an earnings spell (if it is different from the first or last year of the sample window), while in columns (7) and (8), $k = \{1, 2, 3\}$ – to isolate the corresponding first three and last three observations in an incomplete earnings spell. The means and variances of those observations are targeted by matching the regression coefficients from Table 2. We assume that ν_{it} is drawn from distributions with means and variances that depend on whether an individual has missing observations in the interior of a non-consecutive earnings spell, in the beginning of a consecutive earnings spell, or in the end of a consecutive earnings spell (the corresponding means and variances have superscripts m, f, and l, respectively, in Appendix Tables A-7–A-10, which contain full estimation results).[26] Note that allowing for separate shocks in the second and third observations after the missing ones (as well as a few observations prior to the missing ones) is consistent with the evidence of elevated occupation and industry mobility in those periods of individual spells in Table A-6. We further assume that the persistence of the shock ν_{it}, θ, is the same as the persistence of the ϵ_{it} shock because the estimated persistence of

[26] By construction, incomplete consecutive earnings spells have missing observations from t_0 to t_0^i, and/or from T_0^i to T_0, where t_0^i and T_0^i are the first and last years of individual i's incomplete earnings spell.

the transitory component barely changes when we drop outlying observations, which can be verified by comparing the results in columns (3)–(4) with the results in columns (1)–(2).[27] As before, the other moments used for estimating the model are the autocovariance moments in either levels or differences. We rely on the simulated minimum distance method, assuming that all of the innovations are i.i.d. normal, and utilize the optimal weighting matrix estimated by block-bootstrap.[28]

Estimating the extended earnings process results in substantial reduction of the estimated variance of permanent shocks in differences and of transitory shocks in levels.[29] Note that when estimating the extended process, it is essential to account for the first observations around the missing ones, because the other outlying observations will be subsumed in the estimated variance of transitory shocks (as the data allow to observe earnings before and after those earnings records that would help in detecting mean-reversion of those shocks if the first outlying observation is controlled for).[30] This is in contrast to the experiment discussed in the previous section in which it was necessary to drop all three observations surrounding the missing one. Dropping the first observation only leads to the second outlying observation being next to the missing one, and induces the associated biases in the estimated variances of permanent and transitory shocks in levels and differences.

4.4 Simulation

Finally, for completeness, we present a suggestive simulation, consistent with the German data, aimed at replicating the results for the consecutive and non-consecutive samples presented above. We replicate our German unbalanced samples in terms of the number of person-year observations and assume that incomes in the spells starting (ending) in the years other than the first (last) year of the sample are, in addition, affected by a transitory shock, which has a negative mean and high variance as in Table 2.

For the consecutive sample, we assume that persistence of the permanent component is 0.980, the variance of permanent shocks is 0.008, persistence of the transitory component is

[27] We have verified that allowing for a different persistence of the ν shock yields similar results.

[28] Because it is well known that earnings shocks are non-Gaussian, we have also tried estimations which assume that the shocks are drawn from a Student t-distribution with the degrees of freedom estimated from the data by matching kurtosis of the growth in earnings observed in the data. We found that the point estimates in Tables A-7–A-8 were virtually the same (with the estimated degrees of freedom of the Student t-distribution equal to about 4, implying a leptokurtic distribution of the shocks). This is not surprising, since the discrepancy in the estimated variances is the feature of the second moments of the data – and not the higher-order moments – as is highlighted in Eq. (D1)–(L2).

[29] The estimated variance of permanent shocks in growth rates is now smaller than in levels in Table A-7 but if we restrict the persistence of permanent shocks to one in column (5), the estimated variances of permanent and transitory shocks in levels and differences become virtually the same.

[30] It is clear from the precision of our estimates that allowing for the means and variances of rare shocks is not redundant in fitting the data moments; e.g., the quasi-likelihood ratio test's p-values for excluding $\sigma^2_{\nu^f_t}$, $\sigma^2_{\nu^l_t}$, $\mu^l_{\nu^f_t}$, and $\mu^l_{\nu^l_t}$ in the estimation of column (5) Table A-9 are all well below 1%.

0.170, the variance of transitory shocks is 0.010, and the variance of fixed effects is 0.025. These values are similar to the estimates of the transitory component using the moments in growth rates and of the permanent component using the moments in levels in Table 1, columns (1)–(2). We assume that the shocks and fixed effects are drawn from Student t-distributions with four degrees of freedom, since our samples have high excess kurtosis.[31] We take the means and variances of the rare shocks in the first three and last three periods from columns (1) and (5) of Table 2. The results, averaged across 100 simulations, are in Table A-11. Utilizing the full sample results in overestimation of the variance of the permanent (transitory) shock in differences (levels), and it appears that the permanent component is more robustly estimated utilizing the moments in levels while the transitory component is closer to the truth utilizing the moments in differences. Reassuringly, our full-sample estimation results are similar to the data results in Table 1. Dropping the first three and last three observations in an individual's spell aligns the results in levels and differences – see columns (3) and (4) – and correctly recovers the parameters of the underlying earnings process.

For the non-consecutive sample, we assume that persistence of the permanent component is 0.999, the variance of permanent shocks is 0.005, persistence of the transitory component is 0.20, the variance of transitory shocks is 0.01, and the variance of fixed effects is 0.024. This is in line with the estimated permanent component in column (5) and transitory component in column (6) of Table 1. We take the means and variances of rare shocks in the first three and last three periods, and three years before and three years after missing earnings records, from columns (3) and (7) of Table 2, and assume that the shocks follow the moving-average structure of order 1 with the persistence equal to 0.20. The reported results are averages across 100 simulations. The full-sample estimation results in estimates close to the data estimates in Table 1, and recovers fairly well the permanent component using the moments in levels, and the transitory component using the moments in differences – columns (5) and (6) of Table A-11. Dropping the first three and last three observations in an individual's spell, as well as observations surrounding interior missing records, once again aligns the estimates in levels and differences.

These experiments strongly suggest that no other mechanisms but the presence of outlying earnings observations around missing earnings records are responsible for the discrepancy in the estimated variances of the shocks in levels and differences in our administrative earnings data.

[31] Battistin, Blundell, and Lewbel (2009) document the departure of log-income from normality using survey data from the PSID. Assuming normal shocks instead has no impact on our findings. The choice of degrees of freedom for a Student t-distribution of the shocks is consistent with the data; see footnote (28).

5 Implications for Measuring Consumption Insurance

The models with incomplete insurance markets are the workhorses of modern heterogeneous agent macroeconomics. To ensure accurate quantitative analysis using these models and to properly assess implications of various economic shocks and policies, it is essential that these models deliver the correct amount of insurance available to agents. In a seminal contribution, Blundell, Pistaferri, and Preston (2008) have developed the methodology and provided state-of-the-art measures of the insurance against permanent and transitory shocks available to individuals in the data. These estimates have become the key benchmark for assessing the performance of incomplete markets models. Yet, these estimates are based on the abbreviated earnings process in Eq. (1) and therefore ignore the effects of outlying observations at the beginning and end of earnings spells. As we have established above, such estimation leads to important biases in measuring the variances of permanent and transitory shocks. In this section, we show that this translates into potentially sizable biases in measuring consumption insurance against these shocks available to households in the data.

5.1 Insurance Coefficients

In the standard consumption-savings model in which households face permanent and transitory shocks to earnings or income, if the Euler equation holds at equality, consumption growth, Δc_{it}, can be expressed as

$$\Delta c_{it} = \phi_t \xi_{it} + \psi_t \epsilon_{it} + \zeta_{it}, \tag{6}$$

where $1-\phi_t$ is the amount of insurance of permanent shocks, $1-\psi_t$ is the amount of insurance of transitory shocks available at time t, and the random term ζ_{it} represents innovations in consumption independent of the two income components. Blundell, Pistaferri, and Preston (2008) show that the insurance coefficients for permanent and transitory shocks, in the case of a serially uncorrelated transitory component, can be recovered using the following identifying moments:

Permanent insurance:

$$1-\phi_t = 1 - \frac{E[\Delta c_{it}\Delta y_{it-1}] + E[\Delta c_{it}\Delta y_{it}] + E[\Delta c_{it}\Delta y_{it+1}]}{E[\Delta y_{it}\Delta y_{it-1}] + E[\Delta y_{it}\Delta y_{it}] + E[\Delta y_{it}\Delta y_{it+1}]}, \tag{7}$$

Transitory insurance:

$$1-\psi_t = 1 - \frac{E\left[\Delta c_{it}\Delta y_{it+1}\right]}{E\left[\Delta y_{it}\Delta y_{it+1}\right]}, \tag{8}$$

where each expectation (averaging) is taken over all individuals used for estimation of that particular covariance moment in the equations. Since available sample sizes are typically small leading to potentially imprecise estimates of these indentifying moments, the literature relies on a minimum-distance procedure for estimating the model parameters, that utilizes all of the available autocovariance moments in the data.

Note that the parameters ϕ_t and ψ_t reflect the total amount of insurance of permanent and transitory shocks without directly revealing the individual sources and mechanisms of insurance. For example, if y_{it} stands for disposable household income, measured insurance coefficients will reflect insurance due to accumulated assets but also households' advance information about income innovations not available to the econometrician. If y_{it} is measured as gross household earnings, the insurance coefficients would reflect, in part, consumption smoothing due to taxes and transfers. If y_{it} is measured as household head's earnings, the insurance coefficients would also reflect the insurance due to spousal labor supply.

5.2 The Biases in Estimating Insurance Coefficients Due to Presence of Rare Shocks

If rare shocks are present in the data, Eq. (6) can be modified to read as

$$\Delta c_{it} = \phi_t \xi_{it} + \psi_t \epsilon_{it} + \psi_{\text{rare},t} \nu_{it} + \zeta_{it}, \tag{9}$$

where $1 - \psi_{\text{rare},t}$ is the amount of insurance against the rare shock ν_{it}. To the extent that rare shocks are larger in magnitude and are thus harder to insure against, one can expect that $\psi_{\text{rare},t} > \psi_t$.

We now describe the biases associated with ignoring the rare shocks, when they are present in the data, for measuring $1-\phi_t$ and $1-\psi_t$, i.e., the insurance against permanent and transitory shocks. Since the denominators in Eq. (7)–(8) utilize information on earnings data only, we can use our results in Section 2 to characterize the biases in the estimated insurance coefficients.

Consider first an unbalanced sample with consecutive earnings observations such that part of the sample is comprised of individuals who start their incomplete earnings spells at $t > t_0$ while the rest of individuals have nonmissing earnings and consumption data throughout the whole sample period. Notice that the denominator of Eq. (7) is equal to the identifying moment (D1) and will therefore result in an estimate of $\sigma^2_{\xi,t+1} + s_{t,t+1}(\mu^2_\nu + \sigma^2_\nu)$, where μ_ν and σ^2_ν are the mean and variance of the rare shock, respectively. If consumption reacts to the current shocks only (which will be the case when an intertemporal shift of resources is allowed to the extent desired by a household), none of the moments in the numerator of Eq. (7) will be affected, so that the bias in the estimated permanent insurance at $t+1$ will equal $\left(1 - \frac{\phi_{t+1}\sigma^2_{\xi,t+1}}{\sigma^2_{\xi,t+1} + s_{t,t+1}(\mu^2_\nu + \sigma^2_\nu)}\right) - (1 - \phi_{t+1}) = \lambda_{t+1}\phi_{t+1}$, where $\lambda_{t+1} = \frac{s_{t,t+1}(\mu^2_\nu + \sigma^2_\nu)}{s_{t,t+1}(\mu^2_\nu + \sigma^2_\nu) + \sigma^2_{\xi,t+1}}$, and $s_{t,t+1}$ is the share of individuals who started their earnings spells at time $t > t_0$ and have nonmissing earnings and consumption records at t and $t+1$ in the total number of individuals who have nonmissing earnings records at both times t and $t+1$.

Consider next an unbalanced sample with consecutive earnings observations such that part of the sample consists of individuals who end their incomplete earnings spells at $t < T$, while the other individuals have nonmissing earnings and consumption data throughout the whole sample period. In this case, the denominator of Eq. (7) will equal $\sigma^2_{\xi,t} + s_{t-1,t}(\mu^2_\nu + \sigma^2_\nu)$. Since the rare shock is assumed to occur at t and consumption reacts to the current shocks only, the moment $E[\Delta c_{it}\Delta y_{it-1}]$ equals zero, while the moment $E[\Delta c_{it}\Delta y_{it+1}]$ will be identified by averaging over the sample of individuals who have complete earnings spells, and will equal $-\psi_t\sigma^2_{\epsilon_t}$. The moment $E[\Delta c_{it}\Delta y_{it}]$ will, however, be affected by incomplete earnings spells. Averaging over all individuals observed at times $t-1$ and t, the moment will be estimated as $\phi_t\sigma^2_{\xi,t} + \psi_t\sigma^2_{\epsilon,t} + s_{t-1,t}\psi_{\text{rare},t}(\mu^2_\nu + \sigma^2_\mu)$, where $s_{t-1,t}$ is the share of individuals with incomplete earnings spells in the total sample of individuals observed at times $t-1$ and t. Summing up, the bias in the estimated permanent insurance in this case will equal $\left(1 - \frac{\phi_t\sigma^2_{\xi,t} + s_{t-1,t}\psi_{\text{rare},t}(\mu^2_\nu + \sigma^2_\mu)}{\sigma^2_{\xi,t} + s_{t-1,t}(\mu^2_\nu + \sigma^2_\nu)}\right) - (1-\phi_t) = (\phi_t - \psi_{\text{rare},t})\lambda_t$, where $\lambda_t = \frac{s_{t-1,t}(\mu^2_\nu + \sigma^2_\nu)}{s_{t-1,t}(\mu^2_\nu + \sigma^2_\nu) + \sigma^2_{\xi,t}}$. The bias is unambiguously positive and potentially large if $\phi \gg \psi_{\text{rare}}$ (which is likely to hold because permanent shocks are harder to self-insure against), and λ is large (i.e., the mean and/or the variance of the rare shock are larger than the variance of permanent shocks).

Consider now a sample that consists of individuals with missing earnings records at time t, in the interior of the sample period, and individuals with nonmissing earnings and consumption records throughout the sample period. Individuals with missing earnings records at time t will bias the estimated permanent insurance at times $t+2$ and $t-1$. The biases, respectively, are $\phi_{t+2}\lambda_{t+2}$ and $(\phi_{t-1} - \psi_{\text{rare},t-1})\lambda_{t-1}$ with properly defined λ's.

The transitory insurance estimated using Eq. (8) is not systematically biased either for consecutive samples or for not necessarily consecutive samples. This is a direct consequence of our finding in Section 2 that rare shocks do not bias the estimates of the variance of transitory shocks using the moments in differences.

If the transitory component and the rare-shock component are both moving average processes of order 1, the identifying moment (7) should be modified, adding second-order earnings autocovariances to the denominator, and adding the cross-covariances of consumption growth at time t and earnings growth at times $t+2$ and $t-2$ to the numerator. It is straightforward to show that the biases outlined above will change little if θ is close to zero (as is typically found in the data). The identifying moment (8) should also be modified to $1 - \frac{E[\Delta c_{it}\Delta y_{it+2}]}{\frac{1}{\theta}E[\Delta y_{it}\Delta y_{it+2}]}$, and it can be shown following the same arguments as in Section 2 that the presence of rare shocks does not induce a bias in the estimated transitory insurance using that moment.

5.3 Quantitative Implications

5.3.1 Administrative Data

The estimated transmission coefficient for permanent shocks depends on the measurable mean and variance of rare shocks, the share of the sample experiencing those shocks, and the variance of permanent shocks to earnings. It also depends on the true transmission coefficients of permanent and rare transitory shocks to consumption – ϕ and ψ_{rare}, respectively. Unfortunately, these transmission coefficients for households in Denmark or Germany are not known, for the administrative earnings datasets that we have studied above do not include measures of consumption. Consequently, we assume that $\hat{\phi}$ and ψ_{rare} are similar to the values estimated for the U.S. Specifically, following Blundell, Pistaferri, and Preston (2008), we estimate, using the earnings process in Eq. 1, $\hat{\phi} = 0.26$. Setting $\psi_{\text{rare}} = 0.30$ (as estimated using PSID data below), we obtain the true value of ϕ equal to 0.54 and 0.42 for the German consecutive and non-consecutive unbalanced samples, respectively. Assuming instead $\psi_{\text{rare}} = 0.05$, the respective true transmission coefficients for permanent shocks are 0.85 and 0.50. The biases are clearly non-trivial.[32]

5.3.2 U.S. Survey Data from the PSID

We can directly measure all objects of interest and quantitatively evaluate the biases using survey data from the U.S. that contain earnings and consumption measures. To this end, we use the dataset from Blundell, Pistaferri, and Preston (2008), from which we drop (just a few) top-coded male earnings observations. Because our focus in this paper is on earnings, we restrict attention to insurance against the shocks to male earnings.[33] We begin by documenting that the low mean and high variance of earnings observations surrounding the missing ones is also a feature of the PSID data.

Rare shocks in PSID data. Replicating the analysis of administrative datasets in Section 3.4, Table 6 contains the results of a regression of male earnings residuals in the PSID on the dummies for the first and last observations of contiguous earning spells, and observations surrounding missing records. The data span the reporting period 1979–1993, and earnings recorded in, say, year 1979 reflect male earnings received in 1978. In columns (1) and (3), the dummy "Year observed: first" equals 1 if an individual's first earnings record is after 1979, while the dummy "Year observed: last" equals 1 if an individual's last earnings record is prior to 1993. For comparison, in columns (2) and (4), these dummies equal 1 in the first and last

[32] Their magnitudes are very similar using the Danish administrative data.

[33] While this is of interest in itself, the PSID data also allow for a richer study of insurance, e.g., insurance against the shocks to family disposable income. Hryshko and Manovskii (2016) show that a similar bias in the estimated insurance coefficients for permanent family income shocks arises if high volatility and low mean of income observations surrounding the missing ones are not taken into account.

year of the sample window, i.e., 1979 and 1993, respectively. Earnings residuals are about 0.10 log points lower in the few first and last periods (if they differ from the first and last sample years) but are substantially lower in a few periods right after male earnings are missing – column (1). In contrast, earnings residuals are not different from the unconditional mean of zero in the few first and last periods of the sample window – column (2). In columns (3) and (4), we net out the mean effects of outlying observations on the residuals, and then regress squared (net) residuals on the same dummies as in columns (1) and (2), respectively. Squared residuals are lower in the few first sample years and higher in the last sample years due to the well-known increase in male earnings inequality over the life cycle – column (4). The volatility of earnings, however, is much higher in the first and last sample years if individuals' first earnings records are not in the first sample year and last earnings records are not in the last sample year, as can be seen by comparing the first six regressors in columns (3) and (4) of Table 6.[34] We conclude that the patterns of low mean and high variance of earnings residuals around missing earnings records in the PSID data are qualitatively similar to those in the Danish and German administrative data.

The impact of rare shocks and the income process estimates in the PSID. When estimating the male earnings process using the moments in levels and differences, we follow the choices made by Blundell, Pistaferri, and Preston (2008). Since the sample is small, with about 1,750 individuals, the persistence of the permanent component is restricted to unity. The transitory component is modeled as an MA(1) component. As consumption data are expected to help in identifying earnings shocks, both consumption and earnings growth moments as well as the covariance between earnings and consumption growth are used to recover the variance of permanent and transitory shocks to male earnings when using the moments in differences.[35] A diagonal weighting matrix is used to weight the moments in estimation.

The variance of permanent shocks estimated using the moments in levels is 0.017, and 0.071 using the moments in differences. The variance of transitory shocks using the moments in levels is estimated at 0.20, and 0.095 using the moments in differences.

Below we will estimate the complete earnings process that models outlying earnings records. As a first pass for assessing the influence of outlying observations on these estimates, however,

[34]The coefficients on the dummies measure the variances in the respective periods relative to the average variance in the sample overall measured by a constant. For instance, the estimated constant in column (4) is 0.39, so that the variance of residual earnings in the first year for the earnings spells that start in the first sample year equals 0.21 (=0.39 − 0.18), while the estimated constant in column (3) equals 0.34 such that the volatility of earnings in the first year for earnings spells that start later than the first sample year equals 0.43. Similarly, the difference in the volatility of earnings residuals in the last year for the spells that end earlier than the last sample year and spells ending in the last sample year equals 0.34=(0.57 − 0.18) − (0.39 − 0.34).

[35]Consistent with the results in Blundell, Pistaferri, and Preston (2008), the insurance coefficients for permanent and transitory shocks are assumed to be constant over time and life cycle in estimations.

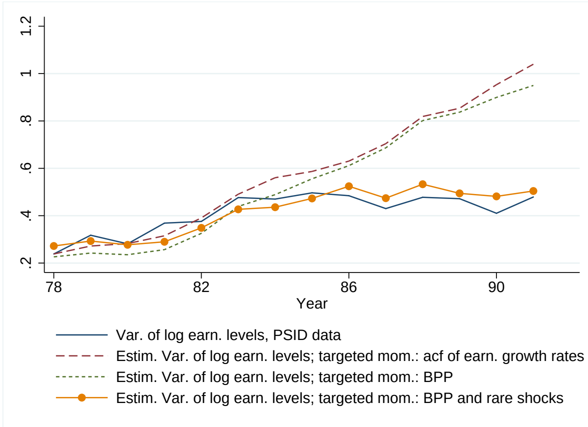
we simply drop the first three and last three earnings observations, if an individual's first record is after 1979 and the last record is prior to 1993, respectively, as well as the three earnings observations before and after missing earnings records. In this sample, the average variance of permanent shocks, using the moments in levels, is estimated at about 0.015, barely changing relative to the estimate for the whole sample. The estimated variance of permanent shocks using the moments in differences is, however, substantially reduced from 0.071 to about 0.022. After dropping outlying observations, the numbers for the variance of transitory shocks are 0.108 and 0.084 when using the moments in levels and growth rates, respectively. While the variance of transitory shocks using the moments in growth rates changes little after dropping outlying observations, the variance of transitory shocks using the moments in levels is cut in half. These patterns in the PSID data are once again qualitatively similar to those in the Danish and German administrative data.

Insurance of the shocks to male earnings in the PSID. Using the full sample, the estimated transmission coefficient for permanent shocks to male earnings is about 0.26, implying insurance of about 74% – see Table 7, Panel A. The value is similar to the one reported in Blundell, Pistaferri, and Preston (2008). After dropping outlying observations, in line with the theoretical bias outlined above, the transmission coefficient for permanent shocks to male earnings rises to 0.53, implying a considerably smaller insurance. [36]

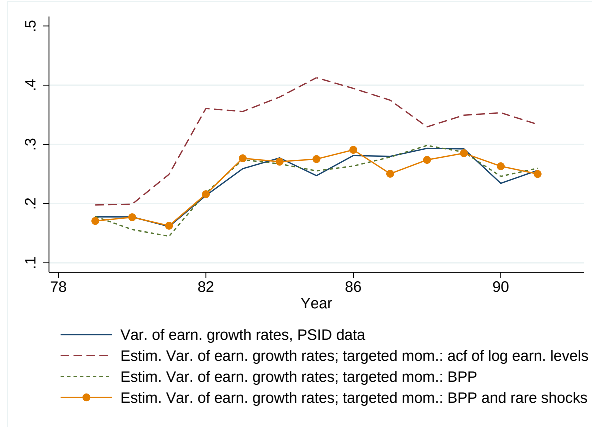
Next, we explicitly recognize the presence of rare shocks in the income process as in Eq. (5) and in the consumption equation (9) and estimate all the parameters of the extended model while retaining all earnings observations. As in Section 4.3, in addition to the standard consumption and income moments we use the regression coefficients in Table 6 to estimate the mean and variance effects of rare shocks. In Panel B of Table 7, we allow for mean and variance effects only for earnings observations right next to a missing record, while in Panel C we allow for the mean and variance effects for three observations around missing records.[37]

[36] We lose about 27% of observations on earnings growth residuals when dropping observations around the missing ones. To verify that the smaller number of usable observations is not the reason for the result, we conducted a Monte Carlo experiment, randomly dropping 27% observations on earnings growth rates. The average transmission coefficient from such an experiment across 1,000 replications was 0.27, virtually the same transmission coefficient as for the full sample.

[37] Specifically, in the full estimation in column (3), in addition to all of the moments in the original BPP estimation we target the regression coefficients in two regressions, with residuals and (net) squared residuals on the left-hand side, and the 19 regressors on the right-hand side: six dummies around interior missing earnings observations, three dummies for the first earnings records if the incomplete earnings spells start later than the first sample year, three dummies for the first earnings records if spells start in the first sample year, three dummies for the last earnings records if the incomplete earnings spells end earlier than the last sample year, three dummies for the last earnings records if earnings spells end in the last sample year, and a constant. We estimated the model by the method of simulated minimum distance, assuming that permanent, transitory, and rare transitory shocks are drawn from normal distributions, and using the diagonal weighting matrix calculated by block-bootstrap. We verified that the simulated method of moments with the assumption of normal permanent and transitory shocks delivers virtually the same parameter estimates as the standard BPP estimation (the results of which are reported in Panel A), which allows for any distributions of permanent and



(a) Variances of log earnings levels



(b) Variances of earnings growth rates

Notes: "Acf" stands for autocovariance function; "BPP" moments include the autocovariance functions of earnings and consumption growth rates, and the cross-covariances between earnings and consumption growth rates; "BPP and rare shocks" moments include, in addition, the regression coefficients reported in Table 6. Male earnings and household nondurable consumption data for 1979–1993 from Blundell, Pistaferri, and Preston (2008) are used in estimations.

FIGURE 2: FIT TO THE MOMENTS OF MALE LOG EARNINGS IN LEVELS AND DIFFERENCES. PSID DATA.

The estimated transmission coefficient for permanent shocks to male earnings is about 0.56 in Panel B and 0.52 in Panel C, not far from the estimated coefficient using the sample with the outlying observations dropped. The insurance against rare transitory shocks, which are larger in magnitude than typical transitory shocks, is estimated at about 70%, while the estimated insurance against typical transitory shocks is not statistically different from 100%.

As discussed above, the bias in measured insurance coefficients caused by neglecting the presence of rare shocks is induced by the misspecification of the earnings process. The consequences of this misspecification are clearly visible in Figure 2. Panel (a) of the Figure plots the variance of log earnings levels in the PSID data over time (solid line) and the fit of various models to these data. The short-dashed line plots the variance implied by the estimates of the model in Blundell, Pistaferri, and Preston (2008) that targets income and consumption growth rates, while the long-dashed line plots the variance implied by targeting income growth rates moments only. Regardless of whether consumption growth moments are used in estimation, by the last sample year the implied variance is about twice as large as the variance in the data. This is the direct consequence of overestimating the variances of permanent shocks when targeting the moments in differences. Targeting the same moments in differences as well as the mean and variances of rare shocks, however, leads to a close match to the (untargeted) variance of earnings levels in the data, as indicated by the line with circles.

Panel (b) of the Figure indicates that both the abbreviated earnings model and the one

transitory shocks.

that explicitly estimates rare shocks provide a good fit to the variance of income growth rates (targeted in estimation). However, the estimation of the abbreviated process targeting the moments in levels substantially overpredicts the observed income growth variances. This is obviously the consequence of overestimating the variance of transitory shocks using the moments in levels.

6 Conclusion

Properties of the earnings process play an important role in various areas of macro and labor economics. Different specifications of this process have been explored in the literature, but the most widely used one is based on decomposing earnings into the sum of persistent and transitory components, where the persistent component is often assumed to follow a random walk. The parameters of such a process can be identified using the moments based on earnings growth rates (first-difference in log earnings) or the moments based on log earnings levels. Historically, the former approach is more common in labor economics, while the latter is more common in the macroeconomics literature. Unfortunately, these two approaches lead to dramatically different estimates of the variances of permanent and transitory components. In particular, using the same set of observations in the data, the variance of the persistent component is typically estimated to be much higher when the moments in growth rates are targeted, while the variance of the transitory component is found to be much higher when the estimation is based on fitting the moments in levels. This has important implications for substantive economic analysis. For example, the earnings process drives the heterogeneity in Bewley-type models with incomplete markets, and the variances of earnings components determine not only economic choices, such as consumption, and savings, but also the optimal design of policies, such as taxes and transfers. Moreover, the standard approach to estimating the amount of insurance that individuals have against permanent and transitory shocks in the data relies on the estimated variances of permanent and transitory components. The uncertainty about the size of these variances translates into uncertainty regarding the right amount of insurance that should be generated by the widely used incomplete markets models, and the associated uncertainty about the results of welfare analyses using those models.

In this paper, we uncovered the features of the data that can quantitatively account for the large difference in the estimates based on earnings growth rates and levels in the administrative data from Denmark and Germany. In particular, we found that earnings are lower on average and more volatile at the start and end of continuous earnings spells. We have shown theoretically that these "outlying" earnings observations, which are either preceded or followed by a missing observation, induce an upward bias in the estimates of the variance of permanent shocks based on the moments in differences and the variance of transitory shocks when estimation is based on the moments in levels. Thus, even when working with very large administrative datasets with highly reliable information, one must remain vigilant because such natural features of the datasets as low mean and high variance of earnings at the start and end of earnings spells can induce extremely large biases in the estimated earnings processes.

While the primary focus of this paper is on estimating earnings processes on large administrative datasets that are becoming central in the literature, the mechanism we describe also applies to survey-based data on earnings or hourly wages. We illustrate the importance of accounting for the high variance and low mean of earnings at the start and end of the earnings spells by replicating the analysis in Blundell, Pistaferri, and Preston (2008) using their PSID male earnings data. We show that not taking these features of the data into account leads to significant upward biases in the estimated amount of insurance against permanent earnings shocks.

These findings have several practical implications for estimation of the earnings process that depend on the objective of the analysis. If one is interested in the properties of permanent and transitory components as well as the detailed analysis of earnings at the start and end of employment spells, one can estimate the extended process in (5) where the mean and the variance of the shocks at the beginning and the end of contiguous earnings histories are readily identified from the mean and the variance of earnings in those periods. Many macro models are too stylized, however, to incorporate explicit treatment of these observations. They use as an input only the permanent/transitory components of the earnings process, as in Eq. (1). The substantive implications of neglecting outlying earnings observations in the macro analysis are probably not very extensive, as the low mean and high variance of earnings in these periods are largely due to time aggregation in the data. It is crucial, however, to obtain the correct estimates of the stochastic properties of the regular permanent and transitory shocks. We find that neglecting outlying observations induces large biases in estimating the variances of these components, which would definitely lead to erroneous substantive conclusions from these models. These components should be estimated using the extended earnings process in (5), although our analysis implies that simpler alternatives are also available. First, we have shown theoretically and verified empirically, that the variance of the transitory shock is estimated with no bias when estimation is based on the moments for earnings growth rates if the rare shock is not serially correlated, and the variance of the permanent shock is unbiased when estimation fits the moments in levels. One could therefore use the estimated permanent component from targeting the moments in levels and the estimated transitory component from targeting the moments in growth rates. An alternative way to proceed would be to estimate the earnings process in Eq. (1) on the data that do not include the observations surrounding the missing ones. As we have shown, this solution recovers the true parameters of this abbreviated process quite well.

References

Abowd, J., and D. Card (1989): "On the Covariance Structure of Earnings and Hours Changes," *Econometrica*, 57(2), 411–445.

Arellano, M., and B. Honoré (2001): "Panel Data Models: Some Recent Developments," in *Handbook of Econometrics, Vol. 5*, ed. by J. J. Heckman, and E. E. Leamer, chap. 53, pp. 3229–3296. Elsevier, Amsterdam.

Attanasio, O., E. Hurst, and L. Pistaferri (2012): "The Evolution of Income, Consumption, and Leisure Inequality in the US, 1980–2010," NBER Working Paper # 17982.

Banks, J., and P. Diamond (2010): "The Base for Direct Taxation," in *Dimensions of Tax Design: The Mirrlees Review*, ed. by The Institute for Fiscal Studies (IFS), chap. 6, pp. 548–648. Oxford University Press for The Institute for Fiscal Studies, Oxford.

Battistin, E., R. Blundell, and A. Lewbel (2009): "Why Is Consumption More Log Normal than Income? Gilbrat's Law Revisited," *Journal of Political Economy*, 117, 1140–1154.

Blundell, R., M. Graber, and M. Mogstad (2015): "Labor Income Dynamics and the Insurance from Taxes, Transfers, and the Family," *Journal of Public Economics*, 127, 58–73.

Blundell, R., L. Pistaferri, and I. Preston (2008): "Consumption Inequality and Partial Insurance," *American Economic Review*, 98(5), 1887–1921.

Browning, M., M. Ejrnæs, and J. Alvarez (2010): "Modelling Income Processes with Lots of Heterogeneity," *Review of Economic Studies*, 77, 1353–1381.

Carroll, C. D. (1997): "Buffer-Stock Saving and the Life Cycle/Permanent Income Hypothesis," *Quarterly Journal of Economics*, 112(1), 1–55.

Castañeda, A., J. Díaz-Giménez, and J.-V. Ríos-Rull (2003): "Accounting for the U.S. Earnings and Wealth Inequality," *Journal of Political Economy*, 111(4), 818–857.

Deaton, A. (1991): "Saving and Liquidity Constraints," *Econometrica*, 59(5), 1121–1248.

DeBacker, J., B. Heim, V. Panousi, S. Ramnath, and I. Vidangos (2013): "Rising Inequality: Transitory or Persistent? New Evidence from a Panel of U.S. Tax Returns," *Brookings Papers on Economic Activity*, pp. 67–142.

Domeij, D., and M. Flodén (2010): "Inequality Trends in Sweden 1978–2004," *Review of Economic Dynamics*, 13(1), 179–208.

Dustmann, C., J. Ludsteck, and U. Schönberg (2009): "Revisiting the German Wage Structure," *Quarterly Journal of Economics*, 124(2), 843–881.

Farhi, E., and I. Werning (2012): "Insurance and Taxation over the Life Cycle," *Review of Economic Studies*, 80, 596–635.

Fitzgerald, J., P. Gottschalk, and R. Moffitt (1998): "An Analysis of Sample Attrition in Panel Data: The Michigan Panel Study of Income Dynamics," *Journal of Human Resources*, 33(2), 251–299.

Friedman, M. (1957): *A Theory of the Consumption Function.* Princeton University Press, Princeton, NJ.

Friedman, M., and S. Kuznets (1954): *Income from Independent Professional Practice.* National Bureau of Economic Research, New York.

Geweke, J., and M. Keane (2000): "An Empirical Analysis of Earnings Dynamics Among Men in the PSID: 1968–1989," *Journal of Econometrics*, 96, 293–356.

Gottschalk, P., and R. A. Moffitt (1994): "The Growth of Earnings Instability in the U.S. Labor Market," *Brookings Papers on Economic Activity*, 2, 217–54.

Guvenen, F. (2009): "An Empirical Investigation of Labor Income Processes," *Review of Economic Dynamics*, 12(1), 58–79.

Guvenen, F., S. Ozcan, and J. Song (2014): "The Nature of Countercyclical Income Risk," *Journal of Political Economy*, 122(3), 621–660.

Heathcote, J. (2005): "Fiscal Policy with Heterogeneous Agents and Incomplete Markets," *Review of Economic Studies*, 72(1), 161–188.

Heathcote, J., F. Perri, and G. L. Violante (2010): "Unequal We Stand: An Empirical Analysis of Economic Inequality in the United States: 1967–2006," *Review of Economic Dynamics*, 13(1), 15–51.

Heathcote, J., K. Storesletten, and G. L. Violante (2010): "The Macroeconomic Implications of Rising Wage Inequality in the United States," *Journal of Political Economy*, 118(4), 681–722.

——— (2014): "Consumption and Labor Supply with Partial Insurance: An Analytical Framework," *American Economic Review*, 104(7), 2075–2126.

Hryshko, D. (2012): "Labor Income Profiles Are Not Heterogeneous: Evidence From Income Growth Rates," *Quantitative Economics*, 3, 177–209.

Hryshko, D., and I. Manovskii (2016): "How Much Consumption Insurance in the U.S.?," mimeo, University of Alberta and University of Pennsylvania.

Hubbard, R. G., J. Skinner, and S. P. Zeldes (1995): "Precautionary Saving and Social Insurance," *Journal of Political Economy*, 103(2), 360–399.

Huggett, M., G. Ventura, and A. Yaron (2011): "Sources of Lifetime Inequality," *American Economic Review*, 101(7), 2923–2954.

Krueger, D., and F. Perri (2006): "Does Income Inequality Lead to Consumption Inequality? Evidence and Theory," *Review of Economic Studies*, 73(1), 163–193.

Krueger, D., F. Perri, L. Pistaferri, and G. L. Violante (2010): "Cross-sectional Facts for Macroeconomists," *Review of Economic Dynamics*, 13(1), 1–14.

Livshits, I., J. MacGee, and M. Tertilt (2007): "Consumer Bankruptcy: A Fresh Start," *American Economic Review*, 97(1), 402–418.

MaCurdy, T. E. (1982): "The Use of Time Series Processes to Model the Error Structure of Earnings in a Longitudinal Data Analysis," *Journal of Econometrics*, 18(1), 83–114.

——— (2007): "A Practitioner's Approach to Estimating Intertemporal Relationships Using Longitudinal Data: Lessons from Applications in Wage Dynamics," in *Handbook of Econometrics, Vol. 6A*, ed. by J. J. Heckman, and E. E. Leamer, chap. 62, pp. 4057–4167. Elsevier, Amsterdam.

Meghir, C., and L. Pistaferri (2004): "Income Variance Dynamics and Heterogeneity," *Econometrica*, 72(1), 1–32.

——— (2011): "Earnings, Consumption, and Life Cycle Choices," in *Handbook of Labor Economics, Vol. 4B*, ed. by D. Card, and O. Ashenfelter, chap. 9, pp. 773–854. Elsevier, Amsterdam.

Moffitt, R. A., and P. Gottschalk (2012): "Trends in the Transitory Variance of Male Earnings: Methods and Evidence," *Journal of Human Resources*, 47(1), 204–236.

Postel-Vinay, F., and H. Turon (2010): "On-the-Job Search, Productivity Shocks, and the Individual Earnings Process," *International Economic Review*, 51(3), 599–629.

Table 1: Estimates of the earnings process in administrative data.

	9 consec.				20 not nec. consec.				Balanced sample			
	German data		Danish data		German data		Danish data		German data		Danish data	
	Levs. (1)	Diffs. (2)	Levs. (3)	Diffs. (4)	Levs. (5)	Diffs. (6)	Levs. (7)	Diffs. (8)	Levs. (9)	Diffs. (10)	Levs. (11)	Diffs. (12)
$\hat{\phi}_p$	0.976	0.992	0.955	0.987	0.999	0.991	0.964	0.982	1	0.998	0.969	0.970
	(0.001)	(0.0008)	(0.0008)	(0.0004)	(0.001)	(0.001)	(0.0007)	(0.0006)	(0.001)	(0.002)	(0.0007)	(0.0009)
$\hat{\sigma}_\xi^2$	0.008	0.019	0.008	0.013	0.0048	0.009	0.007	0.012	0.0031	0.0033	0.005	0.005
	(0.0002)	(0.0003)	(0.0001)	(0.0001)	(0.0001)	(0.0002)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.00004)	(0.00005)
$\hat{\theta}$	0.129	0.153	0.204	0.209	0.119	0.192	0.137	0.217	0.278	0.258	0.212	0.209
	(0.005)	(0.009)	(0.002)	(0.003)	(0.008)	(0.008)	(0.003)	(0.003)	(0.011)	(0.012)	(0.003)	(0.003)
$\hat{\sigma}_\epsilon^2$	0.024	0.009	0.019	0.012	0.016	0.009	0.022	0.013	0.008	0.0078	0.009	0.009
	(0.0003)	(0.0002)	(0.0001)	(0.0001)	(0.0003)	(0.0003)	(0.0002)	(0.0001)	(0.0002)	(0.0002)	(0.0001)	(0.0001)
$\hat{\sigma}_\alpha^2$	0.024	—	0.020	—	0.027	—	0.023	—	0.024	—	0.017	—
	(0.002)	—	(0.0004)	—	(0.002)	—	(0.0004)	—	(0.001)	—	(0.0004)	—
χ^2	1024.43	725.21	7104.42	4527.83	1562.51	1393.16	7002.05	5836.32	1205.84	935.52	6950.05	5855.67
(d.f.)	320	296	346	321	320	296	346	321	320	296	346	321

Notes: The estimated earnings process is: $y_{it} = \alpha_i + p_{it} + \tau_{it}$, where $p_{it+1} = \phi_p p_{it} + \xi_{it+1}$ and $\tau_{it+1} = \epsilon_{it+1} + \theta\epsilon_{it}$. Models are estimated using the optimally weighted minimum distance method. Asymptotic standard errors are in parentheses. German data span the period 1984–2008, while Danish data span the period 1981–2006.

Table 2: Dependent variable: residual earnings and residual earnings squared. Panel regressions.

	Means				Variances			
	9 or more consec.		20 not nec. consec.		9 or more consec.		20 not nec. consec.	
	German data	Danish data	German data	Danish data	German data	Danish data	German data	Danish data
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Year observed: first	–0.57***	–0.48***	–0.65***	–0.47***	0.23***	0.19***	0.29***	0.20***
	(–64.24)	(–119.58)	(–34.49)	(–58.12)	(41.80)	(70.08)	(21.12)	(35.41)
Year observed: second	–0.11***	–0.17***	–0.14***	–0.19***	0.04***	0.08***	0.09***	0.12***
	(–22.57)	(–54.50)	(–11.38)	(–27.24)	(11.54)	(34.16)	(7.91)	(21.04)
Year observed: third	–0.07***	–0.09***	–0.09***	–0.10***	0.02***	0.04***	0.05***	0.05***
	(–16.71)	(–36.26)	(–8.82)	(–16.87)	(7.93)	(21.78)	(5.80)	(12.19)
Year observed: second-to-last	–0.03***	–0.05***	–0.05***	–0.08***	0.01***	0.02***	0.02***	0.03***
	(–8.34)	(–22.93)	(–6.71)	(–15.56)	(5.62)	(11.23)	(4.31)	(7.78)
Year observed: next-to-last	–0.06***	–0.08***	–0.10***	–0.11***	0.03***	0.04***	0.06***	0.06***
	(–13.99)	(–33.78)	(–9.97)	(–20.98)	(10.47)	(20.42)	(7.25)	(12.41)
Year observed: last	–0.43***	–0.28***	–0.47***	–0.32***	0.20***	0.15***	0.25***	0.18***
	(–59.93)	(–85.30)	(–30.92)	(–43.58)	(42.37)	(58.04)	(21.38)	(27.45)
3 years before earn. miss.			–0.03***	–0.03***			0.02***	0.02***
			(–4.59)	(–10.06)			(2.97)	(9.53)
2 years before earn. miss.			–0.05***	–0.04***			0.04***	0.04***
			(–7.76)	(–15.33)			(5.08)	(13.13)
1 year before earn. miss.			–0.27***	–0.26***			0.15***	0.12***
			(–27.16)	(–78.38)			(15.31)	(38.05)
1 year after earn. miss.			–0.39***	–0.43***			0.23***	0.19***
			(–34.56)	(–112.15)			(20.82)	(50.92)
2 years after earn. miss.			–0.15***	–0.14***			0.05***	0.07***
			(–19.29)	(–46.64)			(6.99)	(22.16)
3 years after earn. miss.			–0.13***	–0.11***			0.03***	0.04***
			(–16.97)	(–36.71)			(4.94)	(15.55)
Adj. R sq.	0.126	0.057	0.102	0.079	0.053	0.053	0.050	0.033
No. obs.	379080	2367552	330748	2298429	379080	2367552	330748	2298429
No. indiv.	18130	102825	13635	90668	18130	102825	13635	90668

Notes: German data span the period 1984–2008, while Danish data span the period 1981–2006. The dummies "Year observed: first"–"Year observed: third" are equal to one if an individual's first earnings record is later than in 1984 in German data and 1981 in Danish data, zero otherwise; "Year observed: second-to-last"–"Year observed: last" are equal to one if an individual's last earnings record is earlier than in 2008 in German data and 2006 in Danish data, zero otherwise. Standard errors are clustered by individual; t-statistics are in parentheses. *** significant at the 1% level, ** significant at the 5% level, * significant at the 10% level.

Table 3: Variances of permanent and transitory shocks in the permanent-transitory decomposition of earnings.

	9 consec.				20 not nec. consec.			
	German data		Danish data		German data		Danish data	
	Levs. (1)	Diffs. (2)	Levs. (3)	Diffs. (4)	Levs. (5)	Diffs. (6)	Levs. (7)	Diffs. (8)
Perm. var., full sample, $\hat{\sigma}^2_{\xi}$	0.013	0.024	0.016	0.019	0.0096	0.018	0.013	0.019
Perm. var., outliers, $\hat{\sigma}^2_{\xi,o}$	0.034	0.158	0.053	0.124	–0.009	0.137	–0.004	0.133
Perm. var., net of outliers, $\hat{\sigma}^2_{\xi,n}$	0.010	0.010	0.013	0.013	0.0097	0.0097	0.013	0.013
Trans. var., full sample, $\hat{\sigma}^2_{\epsilon}$	0.020	0.008	0.014	0.009	0.018	0.007	0.019	0.009
Trans. var., outliers, $\hat{\sigma}^2_{\epsilon,o}$	0.143	0.011	0.104	0.022	0.162	0.011	0.173	0.030
Trans. var., net of outliers, $\hat{\sigma}^2_{\epsilon,n}$	0.007	0.007	0.008	0.008	0.007	0.007	0.008	0.008

Notes: The variances are calculated as in Eq. (2)–(4).

Table 4: Estimates of the earnings process in unbalanced samples. German data.

	Full sample		Drop first & last 3 obs.		Model outliers first & last only		Model outliers first & last 3 obs.	
	Levs. (1)	Diffs. (2)	Levs. (3)	Diffs. (4)	Levs. (5)	Diffs. (6)	Levs. (7)	Diffs. (8)
A: 9 or more consec. sample								
$\hat{\phi}_p$	0.976	0.992	0.982	0.994	0.973	1.0	0.975	1.0
	(0.001)	(0.0008)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.002)
$\hat{\sigma}_\xi^2$	0.0078	0.019	0.006	0.005	0.008	0.005	0.007	0.004
	(0.0002)	(0.0003)	(0.0001)	(0.0001)	(0.0002)	(0.0002)	(0.0001)	(0.0002)
$\hat{\theta}$	0.129	0.153	0.197	0.186	0.142	0.141	0.170	0.196
	(0.005)	(0.009)	(0.007)	(0.007)	(0.004)	(0.006)	(0.005)	(0.006)
$\hat{\sigma}_\epsilon^2$	0.024	0.009	0.010	0.009	0.01	0.01	0.01	0.01
	(0.0003)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
$\hat{\sigma}_\alpha^2$	0.024	—	0.019	—	0.022	—	0.025	—
	(0.002)	—	(0.002)	—	(0.001)	—	(0.001)	—
B: 20 not nec. consec. sample								
$\hat{\phi}_p$	0.999	0.991	0.992	0.995	0.999	0.994	0.999	0.995
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.002)
$\hat{\sigma}_\xi^2$	0.0048	0.009	0.0047	0.0046	0.0046	0.0057	0.004	0.005
	(0.0001)	(0.0002)	(0.0001)	(0.0001)	(0.0001)	(0.0002)	(0.0001)	(0.0002)
$\hat{\theta}$	0.119	0.192	0.204	0.190	0.176	0.167	0.201	0.211
	(0.008)	(0.008)	(0.007)	(0.008)	(0.007)	(0.006)	(0.007)	(0.006)
$\hat{\sigma}_\epsilon^2$	0.016	0.009	0.009	0.008	0.0095	0.0096	0.009	0.009
	(0.0003)	(0.0002)	(0.0002)	(0.0002)	(0.0003)	(0.0002)	(0.0003)	(0.0002)
$\hat{\sigma}_\alpha^2$	0.027	—	0.021	—	0.029	—	0.033	—
	(0.002)	—	(0.001)	—	(0.001)	—	(0.001)	—

Notes: Columns (1) and (2) reproduce the corresponding estimates from Table 1. In columns (1)–(4), the estimated earnings process is: $y_{it} = \alpha_i + p_{it} + \tau_{it}$, where $p_{it+1} = \phi_p p_{it} + \xi_{it+1}$ and $\tau_{it+1} = \epsilon_{it+1} + \theta\epsilon_{it}$. In columns (5)–(8), the estimated earnings process is in Eq. (5) in the text. Models are estimated using the optimally weighted minimum distance method. Asymptotic standard errors are in parentheses. German data span the period 1984–2008, while Danish data span the period 1981–2006. Full estimation results for the models in columns (5)–(8) are contained in Tables A-7–A-8.

Table 5: Estimates of the earnings process in unbalanced samples. Danish data.

	Full sample		Drop first & last 3 obs.		Model outliers first & last only		Model outliers first & last 3 obs.	
	Levs. (1)	Diffs. (2)	Levs. (3)	Diffs. (4)	Levs. (5)	Diffs. (6)	Levs. (7)	Diffs. (8)
A: 9 or more consec. sample								
$\hat{\phi}_p$	0.955	0.987	0.957	0.980	0.954	0.985	0.956	0.985
	(0.001)	(0.0004)	(0.001)	(0.001)	(0.0004)	(0.0004)	(0.0004)	(0.0004)
$\hat{\sigma}^2_\xi$	0.008	0.013	0.007	0.007	0.008	0.006	0.008	0.006
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.00004)	(0.00004)	(0.00004)	(0.00004)
$\hat{\theta}$	0.204	0.209	0.225	0.220	0.233	0.190	0.244	0.201
	(0.002)	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.001)	(0.001)
$\hat{\sigma}^2_\epsilon$	0.019	0.012	0.012	0.011	0.013	0.014	0.012	0.013
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.00006)	(0.00006)	(0.0001)	(0.00005)
$\hat{\sigma}^2_\alpha$	0.020	—	0.019	—	0.024	—	0.024	—
	(0.0004)	—	(0.0004)	—	(0.0002)	—	(0.0002)	—
B: 20 not nec. consec. sample								
$\hat{\phi}_p$	0.964	0.982	0.992	0.995	0.963	0.981	0.968	0.982
	(0.001)	(0.001)	(0.001)	(0.001)	(0.0004)	(0.0004)	(0.0004)	(0.004)
$\hat{\sigma}^2_\xi$	0.007	0.012	0.0047	0.0046	0.007	0.008	0.006	0.007
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.00004)	(0.0001)	(0.00003)	(0.0001)
$\hat{\theta}$	0.137	0.217	0.204	0.190	0.219	0.204	0.228	0.233
	(0.003)	(0.003)	(0.007)	(0.008)	(0.002)	(0.002)	(0.002)	(0.002)
$\hat{\sigma}^2_\epsilon$	0.022	0.013	0.009	0.008	0.014	0.013	0.011	0.012
	(0.0001)	(0.0001)	(0.0002)	(0.0002)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
$\hat{\sigma}^2_\alpha$	0.023	—	0.021	—	0.022	—	0.024	—
	(0.0004)	—	(0.001)	—	(0.0002)	—	(0.0002)	—

Notes: Columns (1) and (2) reproduce the corresponding estimates from Table 1. In columns (1)–(4), the estimated earnings process is: $y_{it} = \alpha_i + p_{it} + \tau_{it}$, where $p_{it+1} = \phi_p p_{it} + \xi_{it+1}$ and $\tau_{it+1} = \epsilon_{it+1} + \theta\epsilon_{it}$. In columns (5)–(8), the estimated earnings process is in Eq. (5) in the text. Models are estimated using the optimally weighted minimum distance method. Asymptotic standard errors are in parentheses. German data span the period 1984–2008, while Danish data span the period 1981–2006. Full estimation results for the models in columns (5)–(8) are contained in Tables A-9–A-10.

Table 6: Male Earnings Residuals. PSID data.

Dependent variable	Residuals		Squared residuals	
	(1)	(2)	(3)	(4)
Year observed: first	–0.11***	–0.01	0.09**	–0.18***
	(–4.56)	(–0.43)	(2.15)	(–7.93)
Year observed: second	–0.07***	0.02	–0.01	–0.12***
	(–3.14)	(1.05)	(–0.23)	(–3.90)
Year observed: third	–0.06**	0.01	0.01	–0.17***
	(–2.31)	(0.68)	(0.32)	(–7.86)
Year observed: second-to-last	–0.09***	0.01	0.19***	–0.04*
	(–2.76)	(0.42)	(2.93)	(–1.65)
Year observed: next-to-last	–0.02	0.01	0.23***	0.01
	(–0.74)	(0.72)	(4.09)	(0.20)
Year observed: last	–0.07*	0.00	0.57***	0.18***
	(–1.70)	(0.21)	(6.50)	(3.38)
3 years before earn. miss.	–0.18*	–0.18*	0.45***	0.46***
	(–1.73)	(–1.72)	(2.70)	(2.74)
2 years before earn. miss.	–0.12	–0.11	0.72***	0.69***
	(–1.18)	(–1.09)	(3.21)	(3.11)
1 year before earn. miss.	–0.32**	–0.33**	1.56***	1.57***
	(–2.29)	(–2.34)	(4.62)	(4.62)
1 year after earn. miss.	–1.11***	–1.11***	1.35***	1.40***
	(–9.02)	(–9.00)	(5.17)	(5.40)
2 years after earn. miss.	–0.52***	–0.52***	1.19***	1.23***
	(–3.95)	(–3.92)	(3.24)	(3.31)
3 years after earn. miss.	–0.25**	–0.25**	0.25	0.26
	(–2.10)	(–2.07)	(1.11)	(1.16)
Adj. R sq.	0.039	0.036	0.064	0.057
No. obs.	16496	16496	16496	16496
No. indiv.	1741	1741	1741	1741

Notes: PSID male earnings data span the period 1979–1993. Earnings recorded in year t reflect earnings received in year $t-1$. In columns (1) and (3), the dummies "Year observed: first"–"Year observed: third" are equal to one if an individual's first earnings record is later than in 1979, and are zero otherwise; "Year observed: second-to-last"–"Year observed: last" are equal to one if an individual's last earnings record is earlier than in 1993, and are zero otherwise. In columns (2) and (4), the dummies "Year observed: first"–"Year observed: third" are equal to one if an individual's first earnings record is in 1979, and are zero otherwise; "Year observed: second-to-last"–"Year observed: last" are equal to one if an individual's last earnings record is in 1993, and are zero otherwise. Standard errors are clustered by individual; t-statistics are in parentheses. *** significant at the 1% level, ** significant at the 5% level, * significant at the 10% level.

Table 7: Minimum-Distance Partial Insurance and Variance Estimates.

	Panel A		Panel B		Panel C	
	σ_ξ^2	σ_ϵ^2	σ_ξ^2	σ_ϵ^2	σ_ξ^2	σ_ϵ^2
	(1)	(2)	(3)	(4)	(5)	(6)
1979	0.0215	0.0809	0.00	0.0762	0.00	0.0591
	(0.0105)	(0.0189)	(0.00)	(0.0168)	(0.00)	(0.0168)
1980	0.0215	0.0640	0.00	0.0569	0.00	0.0403
	(0.0105)	(0.0118)	(0.00)	(0.0119)	(0.00)	(0.0120)
1981	0.0215	0.0817	0.00	0.0642	0.00	0.0434
	(0.0105)	(0.0131)	(0.00)	(0.0112)	(0.00)	(0.0122)
1982	0.0634	0.0776	0.0215	0.0690	0.0291	0.0509
	(0.0158)	(0.0151)	(0.0117)	(0.0142)	(0.0120)	(0.0149)
1983	0.0916	0.1075	0.0362	0.0926	0.0510	0.0726
	(0.0252)	(0.0193)	(0.0183)	(0.0188)	(0.0174)	(0.0194)
1984	0.0710	0.0921	0.0141	0.0884	0.0173	0.0717
	(0.0223)	(0.0180)	(0.0149)	(0.0180)	(0.0127)	(0.0195)
1985	0.0819	0.1052	0.0325	0.0893	0.0348	0.0692
	(0.0195)	(0.0167)	(0.0149)	(0.0163)	(0.0146)	(0.0175)
1986	0.0899	0.1088	0.0306	0.0933	0.0248	0.0808
	(0.0194)	(0.0173)	(0.0140)	(0.0168)	(0.0132)	(0.0187)
1987	0.0675	0.0897	0.0500	0.0816	0.00	0.0764
	(0.0258)	(0.0151)	(0.0241)	(0.0149)	(0.00)	(0.0165)
1988	0.0978	0.1319	0.0567	0.1133	0.0002	0.1195
	(0.0316)	(0.0225)	(0.0324)	(0.0225)	(0.0115)	(0.0233)
1989	0.0740	0.0965	0.0719	0.0797	0.0031	0.0651
	(0.0318)	(0.0164)	(0.0346)	(0.0159)	(0.0280)	(0.0170)
1990–1992	0.0527	0.1025	0.0158	0.0940	0.0201	0.0790
	(0.0150)	(0.0121)	(0.0083)	(0.0117)	(0.0082)	(0.0117)
Serial corr. trans. shock	0.0406		0.0067		0.0428	
	(0.0311)		(0.0359)		(0.0357)	
Var. unobs. slope heterog.	0.0137		0.0089		0.0063	
	(0.0037)		(0.0058)		(0.0061)	
ϕ	0.2629		0.5563		0.5224	
(Partial ins. perm. shock)	(0.0549)		(0.1903)		(0.1326)	
ψ	0.0364		–0.0289		–0.0907	
(Partial ins. trans. shock)	(0.0295)		(0.0372)		(0.0670)	
ψ, rare shock	—		0.3556		0.2768	
(Partial ins. rare trans. shock)	—		(0.1021)		(0.0910)	

Notes: Panel A contains the results of the original BPP estimation. In Panel B (C), in addition to all of the moments in the original BPP estimation, we target the regression coefficients in two regressions, with residuals and (net) squared residuals on the left-hand side, and 19 (7) regressors on the right-hand side – six (two) dummies around interior missing earnings observations, three (one) dummies for the first earnings records if incomplete earnings spells start later than the first sample year, three (one) dummies for the first earnings records if spells start in the first sample year, three [one] dummies for the last earnings records if incomplete earnings spells end earlier than the last sample year, three (one) dummies for the last earnings records if earnings spells end in the last sample year, and a constant – and the variance of initial conditions. We estimated the models of Panels B and C by the method of simulated minimum distance, assuming that permanent, transitory, and rare transitory shocks are drawn from normal distributions, and using the diagonal weighting matrix for the moments calculated by block-bootstrap. In all panels, in addition, we estimated the time-varying variances of measurement (and imputation) error in consumption, and the variance of fixed effects in male earnings. In all panels, the variance of permanent shocks in 1979, 1980, and 1981 is restricted to be the same for identification purposes. Standard errors in parentheses.

APPENDIX TABLES

Table A-1: Danish data, 1981–2006. Summary statistics for selected years.

	9 consec.	20 not nec. consec.	Balanced
Number of individuals	102,825	90,668	67,008
Number of observations	2,367,552	2,298,429	1,742,208
Education			
Less than high school	0.227	0.222	0.206
High school degree	0.032	0.031	0.029
Vocational training	0.505	0.521	0.542
Two-year university degree	0.046	0.046	0.047
Bachelors degree	0.125	0.122	0.124
Master or Ph.D.	0.065	0.059	0.051
Earnings			
1985	40,157	40,227	41,383
	(12,831)	(12,889)	(12,278)
1995	48,197	48,444	50,004
	(20,562)	(20,462)	(19,954)
2005	52,656	51,511	53,298
	(26,635)	(26,279)	(25,917)
Spell counts			
Start 1981, end 2006	67,008	80,787	67008
Start after 1981, end 2006	13,439	4,376	0
Start in 1981, end before 2006	17,723	5,210	0
Start after 1981, end before 2006	4,655	295	0
Total	102,825	90,668	67008
Number of spells with 20 or more not nec. consec. observations, by length			
[Proportion of missing observations within spell in square brackets]			
20		1,634 [0.144]	
21		2,009 [0.119]	
22		2,665 [0.096]	
23		3,296 [0.079]	
24		4,486 [0.054]	
25		9,570 [0.030]	
26		67,008 [0.00]	

Notes: Earnings are expressed in 2005 Euros; the standard deviation of earnings is given in parentheses.

Table A-2: German data, 1984–2008. Summary statistics for selected years.

	9 consec.	20 not nec. consec.	Balanced
Number of individuals	18,130	13,635	9,452
Number of observations	379,080	330,748	236,300
Education			
Middle school or no degree	0.05	0.04	0.04
Vocational training	0.72	0.74	0.76
High school degree	0.06	0.05	0.05
College	0.17	0.17	0.15
Earnings			
1985	33,626	33,930	34,559
	(15,876)	(13,323)	(12,881)
1995	45,309	47,180	47,965
	(24,702)	(24,295)	(24,463)
2005	49,121	51,289	52,457
	(36,473)	(37,106)	(37,666)
Spell counts			
Start 1984, end 2008	9,452	11,179	9,452
Start after 1984, end 2008	3,136	1,007	0
Start in 1984, end before 2008	4,463	1,393	0
Start after 1984, end before 2008	1,079	56	0
Total	18,130	13,635	9,452
Number of spells with 20 or more not nec. consec. observations, by length			
[Proportion of missing observations within spell in square brackets]			
20		575 [0.054]	
21		509 [0.054]	
22		623 [0.05]	
23		871 [0.037]	
24		1,605 [0.027]	
25		9,452 [0.00]	

Notes: Earnings are expressed in 2005 Euros; the standard deviation of earnings is given in parentheses.

TABLE A-3: DEPENDENT VARIABLE: RESIDUAL EARNINGS AND RESIDUAL EARNINGS SQUARED. PANEL REGRESSIONS.

	Means				Variances			
	9 or more consec.		20 not nec. consec.		9 or more consec.		20 not nec. consec.	
	German data	Danish data	German data	Danish data	German data	Danish data	German data	Danish data
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Year observed: first	–0.03***	–0.02***	–0.01***	–0.01***	0.04***	0.04***	0.06***	0.05***
	(–13.35)	(–23.35)	(–2.98)	(–10.62)	(17.44)	(57.41)	(20.36)	(63.72)
Year observed: second	–0.02***	–0.01***	0.00**	0.00***	0.01***	0.03***	0.03***	0.03***
	(–7.68)	(–8.16)	(2.05)	(4.26)	(10.09)	(43.70)	(16.58)	(49.39)
Year observed: third	–0.01***	–0.00***	0.01***	0.01***	0.00***	0.02***	0.01***	0.02***
	(–5.93)	(–0.38)	(5.67)	(5.85)	(3.94)	(33.90)	(10.13)	(39.13)
Year observed: second-to-last	0.02***	0.00***	0.01***	0.01***	0.00*	0.01***	0.02***	0.01***
	(7.98)	(3.26)	(5.29)	(8.22)	(1.71)	(13.57)	(9.63)	(21.67)
Year observed: next-to-last	0.01***	–0.01	0.01***	0.01***	0.00***	0.01***	0.02***	0.02***
	(4.49)	(–1.23)	(2.96)	(6.02)	(3.26)	(18.17)	(10.42)	(26.79)
Year observed: last	–0.00	–0.01***	–0.00*	–0.00***	0.02***	0.02***	0.04***	0.03***
	(–0.55)	(–7.04)	(–1.89)	(–3.06)	(10.74)	(24.21)	(16.30)	(32.67)
3 years before earn. miss.			–0.03***	–0.02***			0.02***	0.02***
			(–4.20)	(–9.08)			(2.94)	(9.84)
2 years before earn. miss.			–0.06***	–0.04***			0.04***	0.04***
			(–7.73)	(–16.00)			(5.36)	(14.46)
1 year before earn. miss.			–0.27***	–0.27***			0.15***	0.12***
			(–27.09)	(–80.08)			(14.75)	(38.21)
1 year after earn. miss.			–0.40***	–0.43***			0.24***	0.20***
			(–34.52)	(–111.93)			(20.60)	(51.48)
2 years after earn. miss.			–0.15***	–0.14***			0.06***	0.07***
			(–18.90)	(–46.20)			(7.21)	(22.69)
3 years after earn. miss.			–0.12***	–0.11***			0.04***	0.04***
			(–16.40)	(–36.11)			(5.13)	(16.26)
Adj. R sq.	0.001	0.001	0.051	0.064	0.002	0.002	0.029	0.032
No. obs.	379080	2367552	330748	2298429	379080	2367552	330748	2298429
No. indiv.	18130	102825	13635	90668	18130	102825	13635	90668

Notes: German data span the period 1984–2008, while Danish data span the period 1981–2006. The dummies "Year observed: first"–"Year observed: third" are equal to one if an individual's first earnings record is in 1984 in German data and 1981 in Danish data, zero otherwise; "Year observed: second-to-last"–"Year observed: last" are equal to one if an individual's last earnings record is in 2008 in German data and 2006 in Danish data, zero otherwise. Standard errors are clustered by individual; t-statistics are in parentheses. *** significant at the 1% level, ** significant at the 5% level, * significant at the 10% level.

Table A-4: Dependent variable: the incidence of missing earnings observation; linear probability model. 20 not nec. consec. observations.

	German data					Danish data				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Earn. growth $t-4$ to $t-3$					–0.43***					–0.58***
					(–3.34)					(–10.65)
Earn. growth $t-3$ to $t-2$				–1.02***	–1.16***				–1.06***	–1.23***
				(–6.49)	(–7.04)				(–16.34)	(–18.11)
Earn. growth $t-2$ to $t-1$	–3.21***		–3.10***	–3.32***	–3.41***	–3.15***		–3.01***	–3.25***	–3.33***
	(–12.24)		(–12.14)	(–12.18)	(–12.20)	(–32.02)		(–31.24)	(–32.23)	(–32.57)
Earn. growth $t+1$ to $t+2$		3.58***	3.49***	3.88***	4.03***		4.38***	4.28***	4.77***	4.98***
		(12.67)	(12.60)	(12.80)	(12.96)		(38.88)	(38.54)	(40.48)	(41.34)
Earn. growth $t+2$ to $t+3$				1.32***	1.63***				1.89***	2.29***
				(7.81)	(8.80)				(24.08)	(27.41)
Earn. growth $t+3$ to $t+4$					0.85***					1.27***
					(6.45)					(19.03)
Adj. R sq.	0.005	0.007	0.012	0.013	0.014	0.007	0.012	0.017	0.020	0.021
No. obs.	210641	210641	210641	210641	210641	1486308	1486308	1486308	1486308	1486308
No. indiv.	13635	13635	13635	13635	13635	90584	90584	90584	90584	90584

Notes: Education dummy variables and age are also included in the regressions. Standard errors are clustered by individual; t-statistics are in parentheses. *** significant at the 1% level, ** significant at the 5% level, * significant at the 10% level.

Table A-5: Earnings, wage, and days worked residuals. German data.

	9 consec.			20 not nec. consec.		
	Earn. (1)	Days (2)	Daily Wages (3)	Earn. (4)	Days (5)	Daily Wages (6)
Year observed: first	–0.57***	–0.43***	–0.14***	–0.67***	–0.49***	–0.17***
	(–64.24)	(–57.32)	(–32.99)	(–35.59)	(–31.91)	(–16.48)
Year observed: second	–0.11***	–0.03***	–0.09***	–0.16***	–0.03***	–0.11***
	(–22.57)	(–8.79)	(–23.34)	(–12.50)	(–4.37)	(–11.56)
Year observed: third	–0.07***	–0.02***	–0.05***	–0.11***	–0.02***	–0.08***
	(–16.71)	(–6.76)	(–16.48)	(–10.04)	(–3.08)	(–8.86)
Year observed: second-to-last	–0.03***	–0.02***	–0.01***	–0.05***	–0.02***	–0.03***
	(–8.34)	(–6.66)	(–5.48)	(–6.59)	(–3.19)	(–6.52)
Year observed: next-to-last	–0.06***	–0.03***	–0.03***	–0.10***	–0.04***	–0.06***
	(–13.99)	(–10.61)	(–9.93)	(–10.28)	(–5.66)	(–9.60)
Year observed: last	–0.43***	–0.38***	–0.05***	–0.48***	–0.38***	–0.09***
	(–59.93)	(–59.94)	(–15.03)	(–31.25)	(–30.14)	(–11.49)
3 years before earn. miss.				–0.03***	–0.03***	–0.00
				(–4.19)	(–6.64)	(–0.22)
2 years before earn. miss.				–0.05***	–0.04***	–0.01***
				(–7.74)	(–8.09)	(–2.91)
1 year before earn. miss.				–0.27***	–0.23***	–0.04***
				(–27.10)	(–27.08)	(–8.23)
1 year after earn. miss.				–0.39***	–0.27***	–0.12***
				(–34.42)	(–29.88)	(–23.85)
2 years after earn. miss.				–0.15***	–0.05***	–0.10***
				(–19.07)	(–9.39)	(–20.62)
3 years after earn. miss.				–0.12***	–0.03***	–0.09***
				(–16.42)	(–7.26)	(–17.54)
Adj. R sq.	0.126	0.193	0.013	0.103	0.150	0.021
No. obs.	379080	379080	379080	330748	330748	330748
No. indiv.	18130	18130	18130	13635	13635	13635

Notes: The dummies "Year observed: first"–"Year observed: third" are equal to one if an individual's first earnings record is later than in 1984, zero otherwise; "Year observed: second-to-last"–"Year observed: last" are equal to one if an individual's last earnings record is earlier than in 2008, zero otherwise. Standard errors are clustered by individual; t-statistics are in parentheses. *** significant at the 1% level, ** significant at the 5% level, * significant at the 10% level.

Table A-6: Spell years, unemployment and job mobility. German data.

	9 consec.			20 not nec. consec.		
	Chg. occ. (1)	Chg. ind. (2)	Unemp. (3)	Chg. occ. (4)	Chg. ind. (5)	Unemp. (6)
Year obs.: first	31.05***	31.41***	20.45***	32.40***	32.02***	17.58***
	(41.71)	(42.57)	(31.54)	(21.46)	(21.61)	(13.96)
Year obs.: second	10.05***	9.52***	6.26***	10.18***	8.66***	4.89***
	(17.89)	(18.14)	(14.72)	(8.64)	(8.19)	(6.07)
Year obs.: third	6.01***	5.52***	4.98***	5.64***	4.67***	1.89***
	(12.36)	(12.44)	(12.74)	(5.49)	(5.19)	(2.94)
Year obs: second-to-last	2.18***	2.61***	3.21***	1.40**	3.83***	2.57***
	(6.82)	(8.47)	(11.19)	(2.52)	(5.92)	(4.60)
Year obs.: next-to-last	3.12***	3.27***	3.87***	2.36***	4.41***	3.42***
	(9.19)	(10.17)	(12.60)	(3.91)	(6.61)	(5.55)
Year obs.: last	3.93***	3.71***	24.20***	5.57***	4.75***	23.22***
	(11.06)	(11.20)	(41.08)	(7.57)	(6.96)	(20.15)
3 years before miss.				1.94***	1.72***	3.69***
				(3.62)	(3.50)	(7.09)
2 years before miss.				3.24***	2.91***	4.38***
				(5.37)	(5.45)	(7.71)
1 year before miss.				3.98***	3.69***	22.81***
				(6.34)	(6.37)	(25.55)
1 year after miss.				33.66***	30.83***	17.55***
				(33.82)	(32.37)	(21.26)
2 years after miss.				7.57***	7.78***	5.26***
				(11.65)	(12.30)	(9.45)
3 years after miss.				4.33***	4.30***	3.86***
				(7.57)	(7.93)	(7.53)
Adj. R sq.	0.033	0.036	0.048	0.039	0.042	0.052
No. obs.	378537	378567	379080	330263	330368	330748
No. indiv.	18129	18127	18130	13635	13634	13635

Notes: "Chg. occ." ("Chg. ind.") equals 100 if an individual had changed occupation (industry) between the current year and the last year he had been observed in the sample, zero otherwise. "Unemp." equals 100 if an individual had been unemployed in the current year, zero otherwise. The dummies "Year observed: first"–"Year observed: third" are equal to one if an individual's first earnings record is later than in 1984, zero otherwise; "Year observed: second-to-last"–"Year observed: last" are equal to one if an individual's last earnings record is earlier than in 2008, zero otherwise. We also control for the full set of age dummies in the regressions, and for the dummies that equal one in the first three years of individual spells starting in the beginning of the sample, and the dummies that equal one in the last three years of individual spells ending in the last sample year.

Table A-7: Estimates of the earnings process in unbalanced samples. 9 or more consec. sample. German data.

	Full sample		Drop first & last 3 obs.		Model outliers first & last obs. only		Model outliers first & last 3 obs.	
	Levs. (1)	Diffs. (2)	Levs. (3)	Diffs. (4)	Levs. (5)	Diffs. (6)	Levs. (7)	Diffs. (8)
$\hat{\phi}_p$	0.976	0.992	0.982	0.994	0.973	1.0	0.975	1.0
	(0.001)	(0.0008)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.002)
$\hat{\sigma}^2_\xi$	0.0078	0.019	0.006	0.005	0.008	0.005	0.007	0.004
	(0.0002)	(0.0003)	(0.0001)	(0.0001)	(0.0002)	(0.0002)	(0.0001)	(0.0002)
$\hat{\theta}$	0.129	0.153	0.197	0.186	0.142	0.141	0.170	0.196
	(0.005)	(0.009)	(0.007)	(0.007)	(0.004)	(0.006)	(0.005)	(0.006)
$\hat{\sigma}^2_\epsilon$	0.024	0.009	0.010	0.009	0.01	0.01	0.01	0.01
	(0.0003)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
$\hat{\sigma}^2_\alpha$	0.024	—	0.019	—	0.022	—	0.025	—
	(0.002)	—	(0.002)	—	(0.001)	—	(0.001)	—
$\hat{\sigma}^2_{\nu^f_t}$					0.25	0.23	0.26	0.24
					(0.005)	(0.005)	(0.005)	(0.005)
$\hat{\sigma}^2_{\nu^f_{t+1}}$							0.022	0.031
							(0.003)	(0.003)
$\hat{\sigma}^2_{\nu^f_{t+2}}$							0.010	0.024
							(0.002)	(0.003)
$\hat{\sigma}^2_{\nu^l_t}$					0.21	0.21	0.22	0.21
					(0.004)	(0.004)	(0.004)	(0.004)
$\hat{\sigma}^2_{\nu^l_{t-1}}$							0.025	0.028
							(0.003)	(0.003)
$\hat{\sigma}^2_{\nu^l_{t-2}}$							0.013	0.012
							(0.002)	(0.002)
$\hat{\mu}_{\nu^f_t}$					–0.55	–0.56	–0.57	–0.58
					(0.007)	(0.007)	(0.007)	(0.007)
$\hat{\mu}_{\nu^f_{t+1}}$							–0.02	–0.01
							(0.004)	(0.005)
$\hat{\mu}_{\nu^f_{t+2}}$							–0.08	–0.09
							(0.003)	(0.003)
$\hat{\mu}_{\nu^l_t}$					–0.41	–0.42	–0.42	–0.42
					(0.005)	(0.006)	(0.005)	(0.005)
$\hat{\mu}_{\nu^l_{t-1}}$							–0.057	–0.062
							(0.003)	(0.003)
$\hat{\mu}_{\nu^l_{t-2}}$							–0.04	–0.04
							(0.003)	(0.003)

Notes: In columns (1)–(4), the estimated earnings process is: $y_{it} = \alpha_i + p_{it} + \tau_{it}$, where $p_{it+1} = \phi_p p_{it} + \xi_{it+1}$ and $\tau_{it+1} = \epsilon_{it+1} + \theta\epsilon_{it}$. In columns (5)–(8), the estimated earnings process is in Eq. (5) in the text. Models are estimated using the optimally weighted minimum distance method. Asymptotic standard errors are in parentheses. German data span the period 1984–2008, while Danish data span the period 1981–2006.

Table A-8: Estimates of the earnings process in unbalanced samples. 20 not consec. consec. sample. German data.

	Full sample		Drop first & last 3 obs.		Model outliers first & last obs. only		Model outliers first & last 3 obs.	
	Levs.	Diffs.	Levs.	Diffs.	Levs.	Diffs.	Levs.	Diffs.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\hat{\phi}_p$	0.999	0.991	0.992	0.995	0.999	0.994	0.999	0.995
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.002)
$\hat{\sigma}^2_{\xi}$	0.0048	0.009	0.0047	0.0046	0.0046	0.0057	0.004	0.005
	(0.0001)	(0.0002)	(0.0001)	(0.0001)	(0.0001)	(0.0002)	(0.0001)	(0.0002)
$\hat{\theta}$	0.119	0.192	0.204	0.190	0.176	0.167	0.201	0.211
	(0.008)	(0.008)	(0.007)	(0.008)	(0.007)	(0.006)	(0.007)	(0.006)
$\hat{\sigma}^2_{\epsilon}$	0.016	0.009	0.009	0.008	0.0095	0.0096	0.009	0.009
	(0.0003)	(0.0002)	(0.0002)	(0.0002)	(0.0003)	(0.0002)	(0.0003)	(0.0002)
$\hat{\sigma}^2_{\alpha}$	0.027	—	0.021	—	0.029	—	0.033	—
	(0.002)	—	(0.001)	—	(0.001)	—	(0.001)	—
$\hat{\mu}_{\nu^f_t}$					–0.55	–0.57	–0.62	–0.62
					(0.014)	(0.014)	(0.014)	(0.015)
$\hat{\mu}_{\nu^f_{t+1}}$							0.04	0.03
							(0.008)	(0.008)
$\hat{\mu}_{\nu^f_{t+2}}$							–0.10	–0.11
							(0.007)	(0.007)
$\hat{\mu}_{\nu^l_t}$					–0.33	–0.22	–0.17	–0.13
					(0.011)	(0.01)	(0.008)	(0.007)
$\hat{\mu}_{\nu^l_{t-1}}$							0.023	0.019
							(0.006)	(0.006)
$\hat{\mu}_{\nu^l_{t-2}}$							0.004	–0.005
							(0.005)	(0.006)
$\hat{\sigma}^2_{\nu^f_t}$					0.28	0.31	0.27	0.30
					(0.01)	(0.01)	(0.01)	(0.01)
$\hat{\sigma}^2_{\nu^f_{t+1}}$							0.024	0.045
							(0.008)	(0.009)
$\hat{\sigma}^2_{\nu^f_{t+2}}$							0.040	0.023
							(0.006)	(0.006)
$\hat{\sigma}^2_{\nu^l_t}$					0.14	0.03	0.13	0.08
					(0.009)	(0.006)	(0.006)	(0.005)
$\hat{\sigma}^2_{\nu^l_{t-1}}$							0.016	0.04
							(0.006)	(0.006)

Table A-8 – continued from previous page

	Full sample		Drop first & last 3 obs.		Model outliers first & last obs. only		Model outliers first & last 3 obs.	
	Levs.	Diffs.	Levs.	Diffs.	Levs.	Diffs.	Levs.	Diffs.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\hat{\sigma}^2_{\nu^l_{t-2}}$							0.00	0.02
							(0.00)	(0.004)
$\hat{\mu}_{\nu^m_{t+1}}$					–0.36	–0.33	–0.33	–0.36
					(0.008)	(0.007)	(0.008)	(0.009)
$\hat{\mu}_{\nu^m_{t+2}}$							–0.09	–0.12
							(0.007)	(0.007)
$\hat{\mu}_{\nu^m_{t+3}}$							–0.11	–0.14
							(0.005)	(0.006)
$\hat{\mu}_{\nu^m_{t-1}}$					–0.24	–0.21	–0.19	–0.21
					(0.007)	(0.008)	(0.007)	(0.008)
$\hat{\mu}_{\nu^m_{t-2}}$							–0.04	–0.06
							(0.006)	(0.007)
$\hat{\mu}_{\nu^m_{t-3}}$							–0.01	–0.05
							(0.006)	(0.007)
$\hat{\sigma}^2_{\nu^m_{t+1}}$					0.20	0.13	0.17	0.12
					(0.008)	(0.008)	(0.008)	(0.008)
$\hat{\sigma}^2_{\nu^m_{t+2}}$							0.05	0.015
							(0.007)	(0.007)
$\hat{\sigma}^2_{\nu^m_{t+3}}$							0.04	0.04
							(0.006)	(0.006)
$\hat{\sigma}^2_{\nu^m_{t-1}}$					0.11	0.07	0.09	0.09
					(0.007)	(0.007)	(0.006)	(0.007)
$\hat{\sigma}^2_{\nu^m_{t-2}}$							0.05	0.036
							(0.007)	(0.007)
$\hat{\sigma}^2_{\nu^m_{t-3}}$							0.012	0.03
							(0.005)	(0.006)

Notes: In columns (1)–(4), the estimated earnings process is: $y_{it} = \alpha_i + p_{it} + \tau_{it}$, where $p_{it+1} = \phi_p p_{it} + \xi_{it+1}$ and $\tau_{it+1} = \epsilon_{it+1} + \theta\epsilon_{it}$. In columns (5)–(8), the estimated earnings process is in Eq. (5) in the text. Models are estimated using the optimally weighted minimum distance method. Asymptotic standard errors are in parentheses. German data span the period 1984–2008, while Danish data span the period 1981–2006.

Table A-9: Estimates of the earnings process in unbalanced samples. 9 or more consec. sample. Danish data.

	Full sample		Drop first & last 3 obs.		Model outliers first & last obs. only		Model outliers first & last 3 obs.	
	Levs. (1)	Diffs. (2)	Levs. (3)	Diffs. (4)	Levs. (5)	Diffs. (6)	Levs. (7)	Diffs. (8)
$\hat{\phi}_p$	0.955	0.987	0.957	0.980	0.954	0.985	0.956	0.985
	(0.001)	(0.0004)	(0.001)	(0.001)	(0.0004)	(0.0004)	(0.0004)	(0.0004)
$\hat{\sigma}^2_\xi$	0.008	0.013	0.007	0.007	0.008	0.006	0.008	0.006
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.00004)	(0.00004)	(0.00004)	(0.00004)
$\hat{\theta}$	0.204	0.209	0.225	0.220	0.233	0.190	0.244	0.201
	(0.002)	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.001)	(0.001)
$\hat{\sigma}^2_\epsilon$	0.019	0.012	0.012	0.011	0.013	0.014	0.012	0.013
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.00006)	(0.00006)	(0.0001)	(0.00005)
$\hat{\sigma}^2_\alpha$	0.020	—	0.019	—	0.024	—	0.024	—
	(0.0004)	—	(0.0004)	—	(0.0002)	—	(0.0002)	—
$\hat{\sigma}^2_{\nu^f_t}$					0.17	0.19	0.17	0.20
					(0.002)	(0.002)	(0.005)	(0.002)
$\hat{\sigma}^2_{\nu^f_{t+1}}$							0.10	0.08
							(0.001)	(0.001)
$\hat{\sigma}^2_{\nu^f_{t+2}}$							0.004	0.029
							(0.001)	(0.001)
$\hat{\sigma}^2_{\nu^l_t}$					0.18	0.15	0.18	0.17
					(0.001)	(0.001)	(0.004)	(0.002)
$\hat{\sigma}^2_{\nu^l_{t-1}}$							0.00	0.02
							(0.001)	(0.001)
$\hat{\sigma}^2_{\nu^l_{t-2}}$							0.017	0.02
							(0.001)	(0.001)
$\hat{\mu}_{\nu^f_t}$					–0.39	–0.47	–0.38	–0.47
					(0.002)	(0.003)	(0.002)	(0.002)
$\hat{\mu}_{\nu^f_{t+1}}$							–0.04	–0.08
							(0.001)	(0.002)
$\hat{\mu}_{\nu^f_{t+2}}$							–0.03	–0.09
							(0.001)	(0.001)
$\hat{\mu}_{\nu^l_t}$					–0.29	–0.30	–0.30	–0.31
					(0.002)	(0.002)	(0.002)	(0.002)
$\hat{\mu}_{\nu^l_{t-1}}$							–0.08	–0.07
							(0.001)	(0.001)
$\hat{\mu}_{\nu^l_{t-2}}$							–0.06	–0.06
							(0.001)	(0.001)

Notes: In columns (1)–(4), the estimated earnings process is: $y_{it} = \alpha_i + p_{it} + \tau_{it}$, where $p_{it+1} = \phi_p p_{it} + \xi_{it+1}$ and $\tau_{it+1} = \epsilon_{it+1} + \theta\epsilon_{it}$. In columns (5)–(8), the estimated earnings process is in Eq. (5) in the text. Models are estimated using the optimally weighted minimum distance method. Asymptotic standard errors are in parentheses. German data span the period 1984–2008, while Danish data span the period 1981–2006.

Table A-10: Estimates of the earnings process in unbalanced samples. 20 not consec. consec. sample. Danish data.

	Full sample		Drop first & last 3 obs.		Model outliers first & last obs. only		Model outliers first & last 3 obs.	
	Levs.	Diffs.	Levs.	Diffs.	Levs.	Diffs.	Levs.	Diffs.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\hat{\phi}_p$	0.964	0.982	0.992	0.995	0.963	0.981	0.968	0.982
	(0.001)	(0.001)	(0.001)	(0.001)	(0.0004)	(0.0004)	(0.0004)	(0.004)
$\hat{\sigma}^2_\xi$	0.007	0.012	0.0047	0.0046	0.007	0.008	0.006	0.007
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.00004)	(0.0001)	(0.00003)	(0.0001)
$\hat{\theta}$	0.137	0.217	0.204	0.190	0.219	0.204	0.228	0.233
	(0.003)	(0.003)	(0.007)	(0.008)	(0.002)	(0.002)	(0.002)	(0.002)
$\hat{\sigma}^2_\epsilon$	0.022	0.013	0.009	0.008	0.014	0.013	0.011	0.012
	(0.0001)	(0.0001)	(0.0002)	(0.0002)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
$\hat{\sigma}^2_\alpha$	0.023	—	0.021	—	0.022	—	0.024	—
	(0.0004)	—	(0.001)	—	(0.0002)	—	(0.0002)	—
$\hat{\mu}_{\nu^f_t}$					–0.43	–0.47	–0.46	–0.40
					(0.004)	(0.005)	(0.004)	(0.005)
$\hat{\mu}_{\nu^f_{t+1}}$							–0.09	–0.08
							(0.004)	(0.004)
$\hat{\mu}_{\nu^f_{t+2}}$							–0.02	–0.10
							(0.003)	(0.003)
$\hat{\mu}_{\nu^l_t}$					–0.31	–0.20	–0.08	–0.08
					(0.004)	(0.004)	(0.002)	(0.003)
$\hat{\mu}_{\nu^l_{t-1}}$							0.03	0.06
							(0.002)	(0.002)
$\hat{\mu}_{\nu^l_{t-2}}$							0.02	0.02
							(0.003)	(0.003)
$\hat{\sigma}^2_{\nu^f_t}$					0.18	0.19	0.19	0.20
					(0.003)	(0.004)	(0.003)	(0.004)
$\hat{\sigma}^2_{\nu^f_{t+1}}$							0.10	0.12
							(0.004)	(0.004)
$\hat{\sigma}^2_{\nu^f_{t+2}}$							0.03	0.05
							(0.002)	(0.002)
$\hat{\sigma}^2_{\nu^l_t}$					0.00	0.00	0.07	0.03
					(0.00)	(0.002)	(0.002)	(0.002)
$\hat{\sigma}^2_{\nu^l_{t-1}}$							0.00	0.00
							(0.00)	(0.00)

Table A-10 – continued from previous page

	Full sample		Drop first & last 3 obs.		Model outliers first & last obs. only		Model outliers first & last 3 obs.	
	Levs.	Diffs.	Levs.	Diffs.	Levs.	Diffs.	Levs.	Diffs.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\hat{\sigma}^2_{\nu^l_{t-2}}$							0.01	0.001
							(0.002)	(0.002)
$\hat{\mu}_{\nu^m_{t+1}}$					–0.49	–0.44	–0.53	–0.47
					(0.002)	(0.002)	(0.002)	(0.002)
$\hat{\mu}_{\nu^m_{t+2}}$							–0.06	–0.08
							(0.002)	(0.002)
$\hat{\mu}_{\nu^m_{t+3}}$							–0.11	–0.08
							(0.002)	(0.002)
$\hat{\mu}_{\nu^m_{t-1}}$					–0.27	–0.26	–0.28	–0.27
					(0.002)	(0.002)	(0.002)	(0.002)
$\hat{\mu}_{\nu^m_{t-2}}$							–0.11	–0.02
							(0.002)	(0.002)
$\hat{\mu}_{\nu^m_{t-3}}$							–0.07	–0.04
							(0.002)	(0.002)
$\hat{\sigma}^2_{\nu^m_{t+1}}$					0.24	0.11	0.25	0.12
					(0.002)	(0.002)	(0.002)	(0.002)
$\hat{\sigma}^2_{\nu^m_{t+2}}$							0.04	0.05
							(0.002)	(0.002)
$\hat{\sigma}^2_{\nu^m_{t+3}}$							0.03	0.01
							(0.002)	(0.002)
$\hat{\sigma}^2_{\nu^m_{t-1}}$					0.14	0.09	0.14	0.09
					(0.002)	(0.002)	(0.002)	(0.002)
$\hat{\sigma}^2_{\nu^m_{t-2}}$							0.11	0.014
							(0.002)	(0.002)
$\hat{\sigma}^2_{\nu^m_{t-3}}$							0.03	0.00
							(0.002)	(0.0003)

Notes: In columns (1)–(4), the estimated earnings process is: $y_{it} = \alpha_i + p_{it} + \tau_{it}$, where $p_{it+1} = \phi_p p_{it} + \xi_{it+1}$ and $\tau_{it+1} = \epsilon_{it+1} + \theta\epsilon_{it}$. In columns (5)–(8), the estimated earnings process is in Eq. (5) in the text. Models are estimated using the optimally weighted minimum distance method. Asymptotic standard errors are in parentheses. German data span the period 1984–2008, while Danish data span the period 1981–2006.

Table A-11: Estimates of the earnings process in unbalanced samples. Simulated "German" data.

	9 consec.				20 not nec. consec.			
	Full sample		Drop		Full sample		Drop	
	Levs. (1)	Diffs. (2)	Levs. (3)	Diffs. (4)	Levs. (5)	Diffs. (6)	Levs. (7)	Diffs. (8)
$\hat{\phi}_p$	0.979	0.988	0.980	0.980	0.997	0.995	0.999	0.999
	(0.001)	(0.001)	(0.0009)	(0.001)	(0.0007)	(0.001)	(0.0006)	(0.0007)
$\hat{\sigma}^2_\xi$	0.008	0.016	0.008	0.008	0.005	0.009	0.005	0.005
	(0.0001)	(0.0006)	(0.0001)	(0.0001)	(0.0001)	(0.0003)	(0.0001)	(0.0001)
$\hat{\theta}$	0.133	0.143	0.170	0.170	0.152	0.189	0.20	0.20
	(0.003)	(0.005)	(0.004)	(0.004)	(0.007)	(0.004)	(0.006)	(0.003)
$\hat{\sigma}^2_\epsilon$	0.018	0.009	0.01	0.01	0.014	0.01	0.01	0.01
	(0.0005)	(0.0001)	(0.0001)	(0.0001)	(0.0003)	(0.0001)	(0.0002)	(0.0001)
$\hat{\sigma}^2_\alpha$	0.025	—	0.025	—	0.024	—	0.024	—
	(0.002)	—	(0.002)	—	(0.002)	—	(0.003)	—

Notes: The true earnings process is in Eq. (5). In columns (1)–(4), $\sigma^2_\alpha = 0.025$, $\phi_p = 0.98$, $\sigma^2_\xi = 0.008$, $\theta = 0.170$, $\sigma^2_\epsilon = 0.01$, while in columns (5)–(8), $\sigma^2_\alpha = 0.025$, $\phi_p = 0.999$, $\sigma^2_\xi = 0.005$, $\phi_\tau = 0.20$, $\sigma^2_\epsilon = 0.01$. In columns (3)–(4) the first 3 (last 3) observations are dropped if an individual's earnings spell starts (ends) later (earlier) than in 1984 (2008); in columns (7) and (8), in addition, three observations before and after missing earnings records are dropped. The results are the averages across 100 simulations. The model is estimated using the optimal weighting minimum distance method. Standard errors, calculated as the standard deviations of the estimates across simulations, are in parentheses.